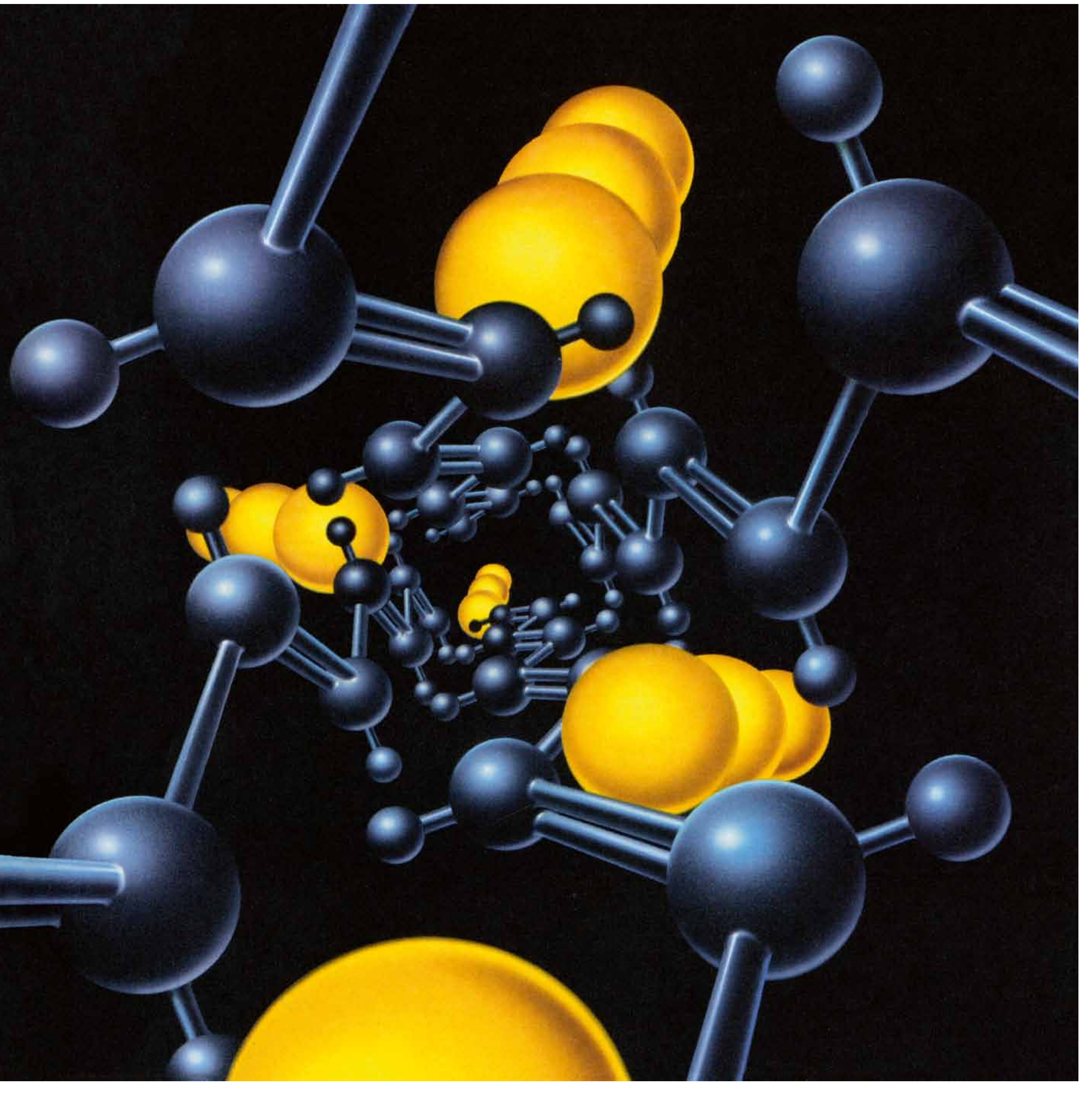
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



PLASTICOS DE CONDUCCION ELECTRICA

Abril 1988
450 PTAS.

INVESTIGACION Y
CIENCIA

Apuntes

En razón de la energía, la humanidad se divide en tres sectores: países paupérrimos si producen menos de 1 kilowatt por persona; países en vías de desarrollo, si producen de 1 a 3 kilowatt por persona, y países desarrollados, si de 3 a 10 kilowatt por persona. Dentro de 50 años el petróleo dejará de constituir una fuente principal de energía. Ante esa perspectiva, la Comunidad Europea ha apostado por la energía de fusión. España ha de aportar a las arcas comunitarias por dicho concepto casi 2000 millones de pesetas anuales.

Los Poríferos o Esponjas sirven de substrato a especies muy diversas de vegetales y animales. A cambio, recaban en tasa de hospedaje, a sus inquilinos, la quitina, un polímero de la N-acetilglucosamina que les es muy necesario y que ellos son incapaces de sintetizar.

Los dos procesos principales de la cicatrización son la migración de las células y su posterior división. Ambos mecanismos vienen instados por los factores de crecimiento. Se acaba de descubrir en ratones de experimentación que un polipéptido de estos, el TGF-B, de origen humano, duplica la eficacia de la cicatrización.

En opinión de algunos científicos ingleses, la intervención ministerial, a través de la aprobación y subvención de determinados proyectos de investigación, atenta contra tres libertades fundamentales: libertad de investigar en temas cuya importancia se desconoce todavía, libertad de cuestionar el saber recibido y libertad de protegerse ante una visión política estrecha del gobierno de turno.

Una suerte de mal blanco está atacando, desde 1983, los arrecifes coralinos del Pacífico. El turno le ha llegado con especial violencia a los celenterados del Atlántico. El emblanquecimiento se debe a la expulsión de las algas que residen en los tejidos blandos del coral y a la pérdida de pigmentos. Las zooxantelas aportan energía y oxígeno a su hospedante, por lo que su pérdida acelera la muerte del coral. Se atribuye ese fenómeno a la elevación de la temperatura del medio.

La insulina es la perla de la bioquímica. Profundamenta estudiada, fue la primera proteína que pudo secuenciarse y sintetizarse; se clonaron los genes que la determinan y se descubrieron mutaciones estructurales. En 1980 la insulina humana constituyó la primera proteína recombinante que se encontraba a disposición de los médicos. Pero esta hormona sigue siendo un maravilloso campo de sorpresas; por citar sólo tres de los últimos avances: se ha hallado que el receptor de la insulina es una quinasa tirosina; se ha clonado el ADN complementario prorreceptor y se ha demostrado que la actividad quinasa del receptor es esencial para la acción insulínica. Por si fuera poco, se han hallado neuropéptidos relacionados con la hormona –que se creía exclusiva de los vertebrados– en los ganglios cerebrales del molusco *Lymnaea stagnalis*. Descubrimiento que, además de dilatar la vida evolutiva de la molécula, plantea el origen nervioso de los péptidos reguladores.

En el extraordinario desarrollo experimentado por la química en las décadas inmediatas a la segunda guerra mundial intervino de manera decisiva la rápida expansión de la espectroscopía de rayos infrarrojos. Gracias a ella, puede medirse cualquier tipo de muestra (líquidos, soluciones, gases, películas, etcétera) en cualquier fase y bajo cualquier condición. Antes de 1939 había un número insignificante de espectrómetros, que se elevó a 700 en 1947, a 3000 en 1958 y a 20.000 en 1969.

La esquistosomiasis es una enfermedad parasitaria que afecta a 200 millones de individuos de todo el mundo y que provoca la muerte anual, según datos de la Organización Mundial de la Salud, de 800.000 personas. El esquistosoma responsable, un platelminto tremátodo, sobrevive en el torrente sanguíneo y escapa a la acción del sistema inmunitario. Para combatirlo se está ensayando una vacuna contra el mismo preparada con proteínas esquistosómicas.

INVESTIGACION Y
CIENCIA

Abril de 1988 Número 139

SECCIONES

COLABORADORES DE ESTE NUMERO

Asesoramiento y traducción:

Santiago Torres Martínez: *Cartografía cromosómica con marcadores de ADN*; Ramón Pascual: *Un defecto en un espejo universal*; Manuel Puigcerver: *Evolución del clima en los planetas terrestres*; Manuel Crespo: *La adaptable zarigüeya*; Amando García Rodríguez: *Plásticos que conducen la electricidad*; Luis Bou: *Ramanujan y el número pi* y *Juegos de ordenador*; Esteban Rodríguez Ocaña: *La Peste Negra*; J. Vilardell: *Taller y laboratorio*.

Ciencia y sociedad:

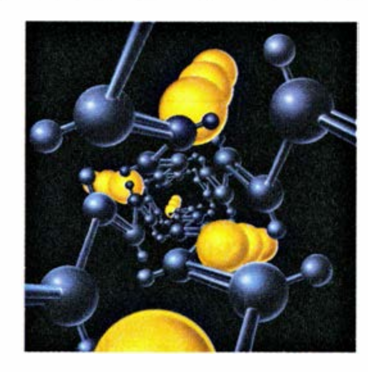
Enrique Julve, José Alcázar y Rosario Cabrero

Libros:

Miguel de Guzmán, Luis Alonso, David Jou y Javier de Lorenzo

PROCEDENCIA DE LAS ILUSTRACIONES

Página	Fuente
13	The Bettmann Archive
14-19	Joan Starwood
20	Edward Bell
23-28	Jerome Kuhl
30	J. C. Carracedo
32-34	J. C. Carracedo y Magda Màrìa
35	J. C. Carracedo
36-37	J. C. Carracedo y Magda Màrìa
38-39	J. C. Carracedo, V. Soler y Magda Màrìa
40-41	J. C. Carracedo
48	National Aeronautics and Space Administration
49	National Aeronautics and Space Administration (*izquierda*), Observatorios Hale (*derecha*)
50-54	George V. Kelvin
55	National Aeronautics and Space Administration
56-57	George V. Kelvin
59	James Kilkelly, por cortesía de la Savannah River Ecology Plant
60-63	Patricia J. Wynne
65	James Kilkelly
66-67	Ian Worpole
68	Instituto de Tecnología de Tokyo
70	James Kilkelly
73	John Moss, por cortesía de The Royal Society de Londres
74	Michael Goodman
75-76	Edward Bell
77	Edward Bell (*arriba*), Laurie Grace (*abajo*)
78	Edward Bell
79	John Moss, por cortesía de The Royal Society de Londres
82-83	Scala/Art Resource
84	The Granger Collection
85-86	Bob Conrad
89	Andrew Christie
90	David Wiseman
91	Richard Dawkins
94	Al Satterwhite, The Image Bank
95-97	Michael Goodman



LA PORTADA presenta un modelo de poliacetileno contaminado, un plástico que, a igualdad de volumen, conduce la electricidad mejor que el cobre (véase "Plásticos que conducen la electricidad", por Richard B. Kaner y Alan G. MacDiarmid, página 64). Los contaminantes, de triyodo (*amarillo*), están situados entre cuatro cadenas de poliacetileno: átomos de carbono unidos a otros de hidrógeno y engarzados entre sí, alternativamente, por medio de enlaces simples y dobles. Como la mayoría de los plásticos, el poliacetileno no contaminado es aislante.

INVESTIGACION Y CIENCIA

DIRECTOR GENERAL Francisco Gracia Guillén
DIRECTOR EDITORIAL José María Valderas Gallardo
REDACCIÓN Carlos Oppenheimer, Josep Maria Farré Josa
PRODUCCIÓN César Redondo Zayas
PROMOCIÓN EXTERIOR Pedro Clotas Cierco
EDITA Prensa Científica, S. A. Calabria, 235-239 - 08029 Barcelona (ESPAÑA)

SCIENTIFIC AMERICAN

DIRECTOR EDITORIAL Jonathan Piel
REDACCIÓN Armand Schwab, Jr. (Subdirector), Timothy Appenzeller, Timothy M. Beardsley, John M. Benditt, Laurie Burnham, David L. Cooke, Jr., Ari W. Epstein, Gregory R. Greenwell, John Horgan, June Kinoshita, Ricki L. Rusting, Karen Wright
DIRECTOR ARTÍSTICO Samuel L. Howard
DIRECTOR DE PRODUCCIÓN Richard Sasso
PRESIDENTE HONORARIO Gerard Piel

SUSCRIPCIONES

Prensa Científica S. A.
Calabria, 235-239
08029 Barcelona (España)
Teléfono 322 05 51 ext. 33-37

Precios de suscripción, en pesetas:

	Un año	Dos años
España	4950	9000
Extranjero	5500	10.000

Ejemplares sueltos:

Ordinario: 450 pesetas
Extraordinario: 575 pesetas

– Todos los precios indicados incluyen el IVA, cuando es aplicable.
– En Canarias, Ceuta y Melilla los precios incluyen el transporte aéreo.
– El precio de los ejemplares atrasados es el mismo que el de los actuales.

DISTRIBUCION

para España:

MIDESA
Carretera de Irún, km. 13,350
(Variante de Fuencarral)
28049 Madrid Tel.652 42 00

para los restantes países:

Editorial Labor, S. A.
Calabria, 235-239 - 08029 Barcelona
Teléfono 322 05 51

PUBLICIDAD

Madrid: Gustavo Martínez Ovín
Menorca, 8, bajo, centro, izquierda.
28009 Madrid
Teléfonos 409 70 45 - 409 70 46
Cataluña:
Paulino Muñoz Victoria
Muntaner, 467-469, 3º 3ª
08021 Barcelona
Teléfono 211 92 45



ISSN 0210-136X Dep. legal: B. 38.999-76

Fotocomposición: Tecfa. Línea Fotocomposición, S.A. Almogàvers, 189 - 08018 Barcelona

Fotocromos reproducidos por Reprostudy, S. A. Buenos Aires, 51 - 08902 Hospitalet (Barcelona)

Gráfica Elzeviriana, S. A. Nápoles, 249 - Tel. 207 40 11 - 08013 Barcelona

Printed in Spain - Impreso en España

Los autores

RAY WHITE y JEAN-MARC LALOUEL ("Cartografía cromosómica con marcadores de ADN") investigan en el Instituto Médico Howard Hugues de la Universidad de Utah. White es también profesor del departamento de genética humana de la Facultad de Medicina y director del centro de investigación sobre fibrosis cística. Doctorado en biología molecular por el Instituto de Tecnología de Massachusetts en 1971, realizó trabajos de postdoctorado sobre bioquímica en la Universidad de Stanford; antes de trasladarse a Salt Lake City formó parte del claustro docente de la Facultad de Medicina de la Universidad de Massachusetts en Worcester. White es uno de los pioneros en cartografía del genoma humano. Lalouel, que pasó su infancia en Africa, enseña genética humana en la Facultad de Medicina de la Universidad de Utah. Licenciado en medicina y doctor en genética cuantitativa por la Universidad de París, donde cursó también estudios de psiquiatría, fue profesor y jefe del departamento de biología humana, incorporándose luego al Instituto Howard Hugues.

ROBERT K. ADAIR ("Un defecto en un espejo universal") es director adjunto de física de altas energías del Laboratorio Nacional Brookhaven. Se recibió de doctor en física en 1951 por la Universidad de Wisconsin, de la que fue instructor hasta 1953. Estuvo seis años en Brookhaven, pasando, en 1959, a la Universidad de Yale. En sus casi treinta años en Yale ha sido profesor, director del departamento de física y director de la división de ciencias físicas. Desde 1978 hasta 1983 dirigió *Physical Review Letters*.

JUAN CARLOS CARRACEDO ("Riesgo volcánico") se licenció en ciencias geológicas por la Universidad Complutense de Madrid (UCM). Después de un periodo de trabajo como investigador contratado en la Universidad de Toronto y en el Centro Espacial de la NASA en Houston, se doctoró en la UCM en 1974. Desde entonces se ha dedicado a la investigación y enseñanza del volcanismo, especialmente del de las islas Canarias donde reside. Responsable científico de la Estación Volcanológica de Canarias desde su creación en 1981, dirige actualmente el Departamento de Volcanología (Estación Volcanológica de Canarias), adscrito al Instituto de Recursos Naturales del CSIC en La Laguna, Tenerife. Formó parte del equipo de volcanólogos que realizó el seguimiento de la erupción del volcán Teneguía -La Palma, 1971- y de la misión científica enviada a Colombia por el gobierno español para el seguimiento de la erupción del volcán Nevado del Ruiz en 1985, adscrito al Comité de Estudios Volcanológicos.

JAMES F. KASTING, OWEN B. TOON y JAMES B. POLLACK ("Evolución del clima en los planetas terrestres") trabajan en el Centro de Investigación Ames de la Administración Nacional de Aeronáutica y del Espacio (NASA). Kasting, que es científico investigador de ese centro, se licenció en física y química por el Harvard College en 1975, doctorándose en ciencias atmosféricas por la Universidad de Michigan en 1979. Becario de postdoctorado en el Centro Atmosférico Nacional, lo fue después en Ames, hasta 1983, cuando pasó a ocupar su cargo actual. Toon ha sido investigador adjunto en Ames desde que se doctoró por Cornell en 1975. Fue uno de los investigadores que propusieron la teoría del invierno nuclear y es el autor de una hipótesis según la cual las nubes que se forman en la estratosfera antártica son de ácido nítrico y contribuyen a la formación del "agujero de ozono". Pollack es científico investigador en Ames. Se licenció en física por la Universidad de California en Berkeley (1962) y doctoró en astronomía por la de Harvard (1965). Realizó trabajos de investigación en el Observatorio Astrofísico Smithsoniano y perteneció a la plantilla del Centro de Radiofísica e Investigación Espacial de Cornell, incorporándose luego a la NASA.

STEVEN N. AUSTAD ("La adaptable zarigüeya") es profesor de biología de la Universidad de Harvard. En 1969, habiéndose licenciado en filología inglesa por la Universidad de California en Los Angeles y cuando trabajaba de amaestrador en Hollywood, se interesó por el comportamiento animal. Cursó entonces la licenciatura en biología por la Universidad estatal de California en Northridge (1976) y se doctoró en ecología etológica por la de Purdue (1981). Fue profesor visitante en Purdue y en la Universidad de Nuevo México; en 1986 se trasladó a Harvard. Austad está casado con una veterinaria y asegura que tiene la "casa llena de animales remendados".

RICHARD B. KANER y ALAN G. MACDIARMID ("Plásticos que conducen la electricidad") trabajaron juntos cuando Kaner, alumno aún, colaboró con el grupo de investigación de MacDiarmid. Hoy Kaner es profesor de química de la Universidad de California de Los Angeles. En 1984 se doctoró en química inorgánica por la Universidad de Pennsylvania; posteriormente amplió estudios en la Universidad de California en Berkeley. Ocupó su plaza actual en 1987. Las investigaciones de Kaner se centran en la química del estado sólido sintético. MacDiarmid enseña química en Pennsylvania. Nacido en Nueva Zelanda, comenzó allí sus estudios y se doctoró en química por la Universidad de Cambridge en 1955. A continuación enseñó en la Universidad de St. Andrews, pasando a Pennsylvania en 1964. Los trabajos de MacDiarmid y sus colaboradores condujeron a la síntesis, contaminación y caracterización del poliacetileno, el primer ejemplo de polímero conductor.

JONATHAN M. BORWEIN y PETER B. BORWEIN ("Ramanujan y el número pi") son hermanos. Jonathan, el mayor, es profesor de matemáticas de la Universidad Dalhousie, Nueva Escocia. Completó el primer ciclo universitario en 1971, en la Universidad de Ontario Occidental, cuyo departamento de matemáticas dirige su padre. Se recibió de doctor en matemáticas por la Universidad de Oxford e inmediatamente ingresó en el claustro docente de Dalhousie. El mayor de los Borwein también ha impartido docencia en la Universidad Carnegie-Mellon y ha investigado en las universidades de Cambridge, Limoges y Montreal. Peter Borwein es profesor de matemáticas de Dalhousie. Comenzó su formación superior en la Universidad de Ontario Occidental (1974) y se doctoró por la de Columbia Británica. El más joven de los Borwein pasó el curso 1979-80 en Oxford y desde entonces es profesor de Dalhousie.

COLIN McEVEDY ("La Peste Negra") es psiquiatra del Hospital San Bernardo, en los aledaños de Londres. Se licenció en medicina en 1955 por la Universidad de Oxford. Mientras sirvió en la fuerza Aérea británica trabajó en el Instituto de Medicina Aeronáutica. Posteriormente se especializó en psiquiatría en el Hospital Maudsley y en el Middlesex. McEvedy, doctor en medicina por Oxford desde 1971, es un amante de la historia, de cuya disciplina ha publicado varios atlas.

Cartas

Señor director,

Como egiptóloga, me sorprendió la fotografía aérea de una cometa del desierto que aparece en "Caza de gacelas en la Siria de la edad de Piedra", de Anthony J. Legge y Peter A. Rowley-Crowy [INVESTIGACIÓN Y CIENCIA, octubre de 1987]. Los antiguos egipcios representaron alrededor del 3100 a. de C., en una paleta votiva que relataba las conquistas del rey Narmer, de la primera dinastía, un motivo de forma casi idéntica. Situado encima de la figura de un enemigo caído, la forma pretende a todas luces identificar la tribu, o las gentes, que representa, y tradicionalmente se había interpretado como algún tipo de forma vegetal. Sin embargo, en 1955, el arqueólogo israelí Yigael Yadin señaló su similitud con la forma de las cometas del desierto vistas en fotografías aéreas. Ninguno de sus ejemplos se parece tanto al signo de Narmer como el de la ilustración del artículo.

Junto al hombre de la cometa yace otro enemigo, identificado por un rectángulo almenado, representación inconfundible del plano de una fortaleza o una ciudad fortificada. Los dos vencidos parecen formar un grupo, separado del resto del dibujo por una gruesa línea horizontal. Yasin propuso la hipótesis, verosímil, de que se trataba de gentes derrotadas en acciones realizadas fuera del valle del Nilo. Dada su familiarización con los restos de cometas del Jordán septentrional, supuso que el conflicto tuvo lugar en ese territorio. Pero esa zona queda muy apartada de Egipto. En el epígrafe que acompaña la ilustración, los autores mencionan la existencia de restos de cometas en el Sinaí, campo de batalla mucho más probable, en particular porque abundan en él los signos de actividad egipcia de finales del cuarto milenio a. de C.



Paleta de Narmer (Museo de Egipto, El Cairo)

Tanto si los extranjeros derrotados del pie de la paleta de Narmer representan territorios, como si corresponden a reinos o tribus, la fortaleza y la cometa los adscriben a tipos de vida muy distintos. Cuando Yadin propuso su hipótesis, la cometa se consideraba un recinto donde los pastores nómadas protegían el ganado. Si, por el contrario, sólo se utilizaban en la matanza en masa (como se describe en el artículo), parecería que los egipcios —que sin duda conocían la función de algo que representaban con tal detalle— pretendían distinguir poblaciones enteramente agrícolas, asentadas en comunidades fortificadas, de gentes que, nómadas o sedentarias, seguían dependiendo en gran medida de la caza.

En el antiguo Egipto era práctica común la caza en el desierto, a menudo recogida en las representaciones gráficas. Esas escenas contienen un elemento simbólico, y a veces deportivo, así como un considerable grado de licencia pictórica. Pero las imágenes muestran con claridad que uno de los métodos de caza, tan importante como el de rastreo y acoso, se servía de un recinto vallado o dotado de red. Nunca se ha retratado enteramente esa construcción. Sin embargo, varias características comunes a todas las representaciones de esos cercados sugieren que se instalaban para grandes batidas, realizadas cada año en alguna estación. De ser así, probablemente el recinto egipcio de caza constituyera una versión reducida de cometa del desierto.

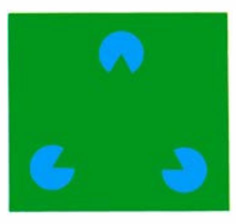
El agotamiento, y en última instancia extinción, de las piezas de caza de los desiertos vecinos al antiguo Egipto se han atribuido a la desecación y a su exterminio. En Egipto, igual que en otras zonas semiáridas del Oriente Próximo, la contribución humana al empobrecimiento de la fauna silvestre ha contado con la mortífera eficacia de las cometas del desierto.

EDNA R. RUSSMANN

Conservador
Departamento de Arte Egipcio
Museo Metropolitano de Arte
Nueva York

Señor director,

De haber sabido que la sección "Taller y laboratorio" del número de marzo iba a tratar de los contornos subjetivos, en mi artículo "Arte, ilusión y sistema visual", publicado en el mismo número, habría incluido la observación siguiente. Según ciertos experimentos que realicé con David Hubel, la impresión de bordes ilusorios no es cognoscitiva. Hemos creado figuras de Kanizsa (imágenes que producen la sensación de que existen bordes donde no los hay) empleando dos colores. Descubrimos que, cuando los colores eran equiluminantes, los bordes ilusorios desaparecían, aunque las porciones no ilusorias de las figuras siguieran mostrándose perfectamente nítidas. Por consiguiente, las tres figuras de la ilustración que sigue, en azul sobre verde, no crearán ningún triángulo si los colores son equiluminantes.



En efecto, en condiciones de equiluminancia se aprecian nítidamente tres círculos a los que se les ha suprimido un sector, pero no se reproduce la impresión de que les tape parcialmente un triángulo. El nivel del sistema visual donde se generan los bordes ilusorios no puede ser suficientemente elevado para considerarlo "cognoscitivo", pues nuestra impresión consciente de las partes no ilusorias de las figuras es igual bajo equiluminancia y sin ella.

La interpretación que damos a ese resultado es que, de las dos subdivisiones principales del sistema visual humano, sólo el sistema "magno", que no distingue el color, tiende a enlazar bordes colineales. Tal facultad determina de ordinario qué bordes de una escena corresponden al mismo objeto, y ello incluso si los bordes están interrumpidos; los bordes ilusorios son meras manifestaciones de ese proceder. Cuando no se da equiluminancia, las dos subdivisiones del sistema visual "ven" los círculos incompletos, pero sólo el sistema magno relaciona los sectores circulares ausentes e indica la presencia de una figura ilusoria. En condiciones de equiluminancia sólo el sistema parvo sigue viendo los contornos.

MARGARET S. LIVINGSTONE

Facultad de Medicina de Harvard
Boston, Massachusetts

Hace...

José M.ª López Piñero

...cuatrocientos cincuenta años

Pedro Sánchez Ciruelo publicó en Salamanca su *Reprovación de las supersticiones y hechizerías*, título fundamental de la literatura antisupersticiosa española del siglo XVI.

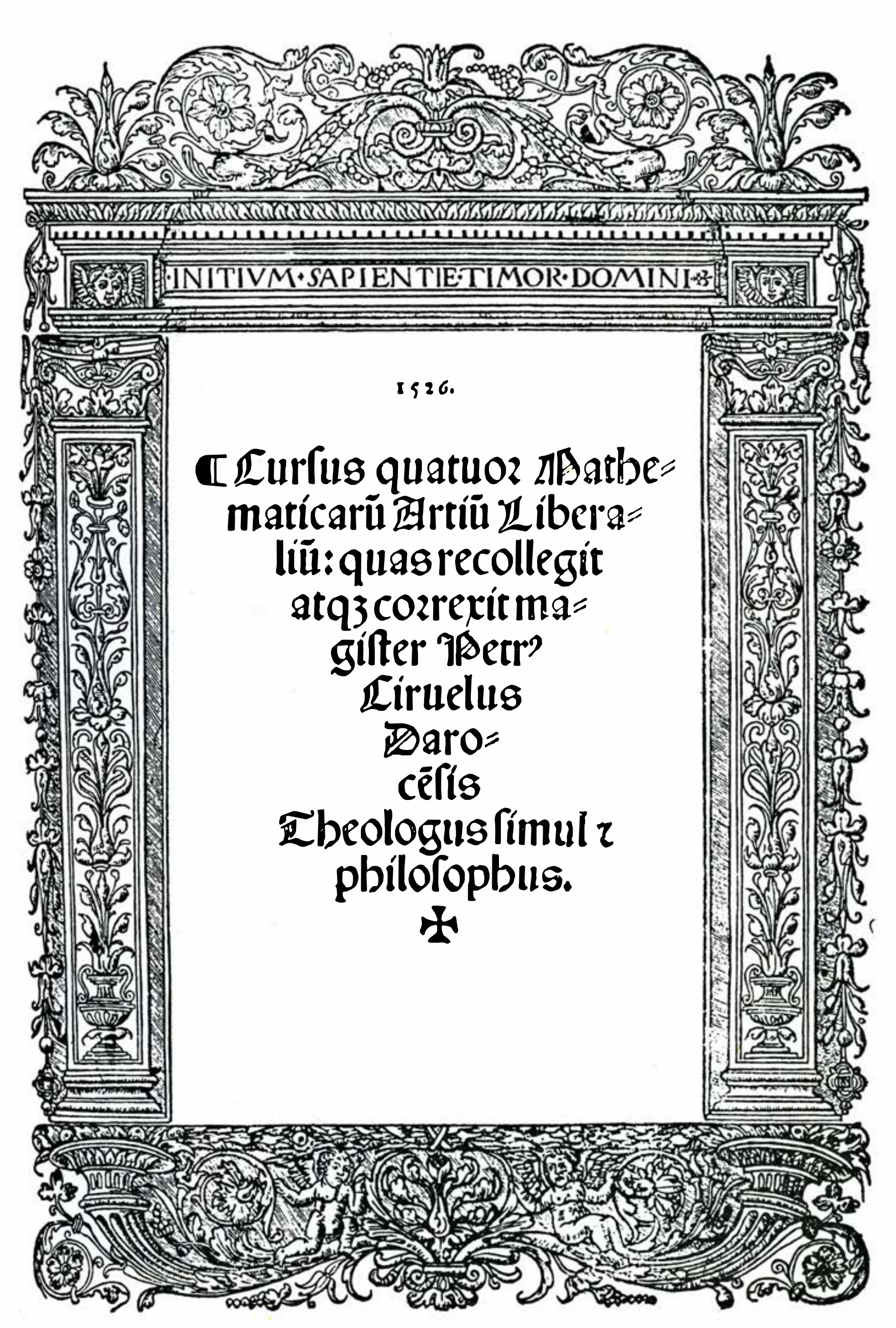
Nacido en Daroca hacia 1470, Ciruelo estudió en la Universidad de Salamanca, donde se formó como matemático bajo el magisterio de Rodrigo Basurto, titular de la cátedra de astronomía. Residió más tarde diez años en París (1492-1502), en cuya Universidad fue discípulo del también aragonés Gaspar Lax y alternó su formación como teólogo con la enseñanza de las matemáticas. De regreso a España, tras una breve estancia en la Universidad de Sigüenza, fue elegido por el cardenal Cisneros para ocupar una cátedra de teología en Alcalá, ciudad en la que permaneció casi un cuarto de siglo. En los últimos diez años de su vida (1538-1548) volvió, sin embargo, a residir en Salamanca. Ciruelo fue autor de una amplia y difundida obra escrita, cuya mayor parte corresponde a diferentes áreas científicas. La estrictamente matemática consta de cuatro títulos, muy representativos tanto por su contenido como por su intención de la línea académica de la disciplina.

Los dos primeros, publicados por vez primera en París el año 1495, son ediciones de la *Arithmetica speculativa* y de la *Geometria speculativa*, de Bradwardine. Anotemos que la edición "bene revisa et correcta" se destina "a los estudiantes de artes y de filosofía aristotélica".

El tercero es el libro titulado *Tractatus arithmetice practice qui dicitur algorismus*, impreso también por vez primera en la capital francesa en las últimas fechas del siglo XV o las primeras del XVI. El adjetivo "práctica" no debe hacer pensar en un tratado de cálculo mercantil o algo parecido. Ciruelo trata en este texto de los enteros y de las fracciones físicas por su interés para los estudiantes de filosofía natural, así como de las fracciones sexagesimales por su utilidad en cosmografía. Por el contrario, excluye explícitamente lo que llama "cuestiones difíciles y curiosas de los comerciantes".

El cuarto, redactado a diferencia de los anteriores durante su estancia en Alcalá y destinado a "servir a tan preclara Universidad", es el *Cursus quatuor mathematicarum artium liberalium*. Su primera edición apareció el año 1516 en dicha ciudad, impresa por el célebre Arnao Guillén de Brocar. Incluye cuatro partes, todas las cuales consisten en textos de otros autores, comentados por Ciruelo y adicionados con diversos fragmentos. La aritmética y la geometría son, de nuevo, las de Bradwardine; la perspectiva, de John Peckham; la música, de Jacques Lefèvre d'Etaples. No puede darse mayor atenimiento a la tradición medieval de las facultades de artes. Ciruelo, no solamente espera que su obra "será fundamento muy estimable para la plena exigencia de ambas filosofías", sino que considera, de acuerdo con San Agustín, que las matemáticas son "necesarias para los teólogos".

Todos estos textos responden a un enfoque prácticamente uniforme, dentro del cual no aparecen aportaciones originales de relieve. Según Lorente y Pérez se reducen a un método de aproximación de las raíces cuadrada y cúbica; a la consideración de los polígonos estrellados como resultado de unir los puntos de división de una circunferencia; a la generalización del teorema de Campano a todos los polígonos estrellados de cualquier orden; y a de-



Portada de la segunda edición del Cursus quatuor mathematicarum artium, *de Ciruelo (Alcalá, 1526).*

mostrar los errores de cuadradores del círculo, subrayando que la cuadratura resultaría posible si se conociera exactamente la razón de la circunferencia al diámetro.

Ciruelo publicó, además, entre otros libros, una edición latina comentada de la *Sphera* de Sacrobosco (1498), varias obras de lógica y astrología, un estudio *De correctione Kalendarii* (1528) y unas *Paradoxae questiones* (1538) que contienen sus ideas sobre mecánica.

Aunque la mayoría de los libros de Ciruelo alcanzaron varias ediciones en España y Francia, el más difundido fue *Reprovación de las supersticiones y hechizerías*, que tuvo nueve reimpresiones hasta mediados del siglo XVI.

El complejo mundo que Ciruelo llamó "supersticiones y hechicerías" comprendía numerosos aspectos puramente creenciales que nada o muy poco tienen que ver con la actividad científica, aunque tengan gran interés desde la perspectiva, por ejemplo, de un historiador de la medicina. Sin embargo, otros aspectos, especialmente en el nivel "erudito" de dicho mundo, llegaron a constituir una auténtica subcultura científica. Los más importantes, sin duda, fueron la astrología judiciaria y la magia negra.

A diferencia de lo que sucedía con la alquimia, la Inquisición los reprimió desde una motivación religiosa, bien fuera por la condición herética de las doctrinas que implicaban o porque supusieran una acción demoníaca. La represión fue dirigida contra sus cultivadores, que aparecen con relativa frecuencia en los procesos inquisitoriales, y también contra los escritos que las defendían. Basta consultar a este respecto los índices de libros prohibidos, particularmente el del inquisidor general Gaspar de Quiroga (1583), cuya "regla novena" está consagrada a su explícita condenación.

Por otra parte, la astrología judiciaria y la magia negra fueron refutadas desde la ortodoxia religiosa y los supuestos de la ciencia académica por las dos tendencias que en la literatura antisupersticiosa de la época ha distinguido Granjel. La primera estuvo integrada por libros en castellano de intención divulgadora, entre los que destacan el *Tratado muy sotil y bien fundado de las supersticiones y hechizerias* (1529), de Martín de Castañega y el de Pedro Ciruelo. La segunda, por obras en latín de ambición doctrinal, las más importantes de las cuales fueron *Adversus fallaces et superstitiosas artes, id est, de Magia* (1603), de Benito Perera, y *Disquisitionum Magicarum Libri Sex* (1624), de Martín del Río.

No vamos a detenernos en el intrincado panorama que, entre otros autores, han estudiado Caro Baroja desde una perspectiva general y Granjel en sus relaciones con la medicina. Será suficiente que intentemos delimitar la astrología judiciaria y la magia negra de la astrología y de la magia permitidas por la ortodoxia religiosa y aceptadas por la comunidad académica.

La astrología, como aplicación de la ciencia de los astros al pronóstico de los sucesos, era una vertiente aplicada de la cosmografía, inseparable del conocimiento teórico en la actividad concreta de cualquiera de sus cultivadores, aunque en abstracto pueda distinguirse de éste. Constituía, por lo tanto, una "lícita y verdadera ciencia". Pedro Ciruelo así lo afirma rotundamente: "Los cielos y estrellas, alterando el aire y la tierra, también alteran a los hombres y a las otras animalias que moran en la tierra y en el aire y en el agua, y ansí los cielos causan en nuestros cuerpos diversas inclinaciones y habilidades a muchas artes y ciencias (...) el verdadero filósofo que conoce las virtudes y propiedades de las estrellas, podrá por ellas conocer los efectos sobredichos en los elementos y en los hombres y árboles, y podrá naturalmente decirlos antes que vengan (...) esta astrología es lícita y verdadera ciencia como la filosofía natural o la medicina".

La astrología judiciaria era, por el contrario, un "arte falaz y supersticiosa" precisamente porque convertía el pronóstico en adivinación determinista, lo que suponía negar el dogma del libre albedrío del hombre e ignorar la azarosidad de muchos sucesos naturales. La distinción la expone con toda claridad la "regla novena" del índice del inquisidor Quiroga: "Se prohiben todos los libros, tractados y escriptos, en la parte que tratan y dan reglas y hacen arte o ciencia para conocer por las estrellas y sus aspectos (...) lo por venir que está en la libertad del hombre y los casos fortuitos que han de acontecer, o que enseñan a responder lo hecho o acontescido en las cosas pasadas libres y ocultas, o lo que sucederá en lo que depende de nuestra libertad, que son las partes de la judiciaria que llaman de nascimientos, interrogaciones y elecciones (...). Pero no por esto se prohiben las partes de la astrología que tocan al conocimiento de los tiempos y sucesos generales del mundo, ni las que enseñan por el nacimiento de cada uno a conocer sus inclinaciones, condiciones y cualidades corporales, ni lo que pertenece a la agricultura y navegación y medicina y las elecciones que cerca de estas cosas naturales se hacen".



Grabado de una de las reediciones del Tractatus de sphera *de Sacrobosco, comentado por Ciruelo (París, 1498).*

La distinción, muy clara en teoría, era más difícil de precisar en la práctica en un caso concreto. De hecho, los sospechosos de cultivar la astrología judiciaria no fueron perseguidos de modo sistemático hasta que Sixto V la condenó enérgicamente en su Bula *Coeli et Terrae* (1585), que no fue publicada en España hasta 1612. Fue entonces cuando la Inquisición procesó, al final de sus vidas, a los tres personajes en los que Caro Baroja ha ejemplificado esta actividad: el valenciano Juan Piquer, formado en "matemáticas" y astrología en la Universidad de Valencia y luego discípulo en Nápoles del propio Gianbattista della Porta; el presbítero, asimismo valenciano, Juan Ramírez; y Luis Rosicler, bordador madrileño cuñado de Lope de Vega, que había asistido como éste a las lecciones de Labaña en la Academia de Matemáticas.

La "magia natural" era otro componente habitual de la ciencia académica del siglo XVI, aunque distara mucho de ser aceptada de modo casi general como la astrología. Estaba basada en la idea neoplatónica de una "simpatía" natural entre todos los seres del universo, que podía ser patente u oculta. Consistía concretamente en el arte de conocer y aprovechar esas relaciones ocultas. Como ha afirmado Marie Boas, era "la forma oculta de la todavía no inventada física experimental". Bajo el influjo directo o indirecto del neoplatonismo, la magia natural fue uno de los supuestos de importantes tendencias del pensamiento y de la ciencia renacentista, desde las que personificamos en Pico della Mirandola, Marsilio Ficino o Paracelso hasta las encabezadas por Fracastoro, Porta, Cardano o Giordano Bruno.

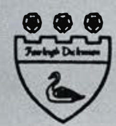
La magia negra, en cambio, suponía la intervención demoníaca. "Magi pos-

PEDRO CIRUELO'S
A TREATISE REPROVING
ALL SUPERSTITIONS AND
FORMS OF WITCHCRAFT

Very Necessary and Useful
for All Good Christians
Zealous for Their Salvation

Translated by
Eugene A. Maio and D'Orsay W. Pearson

Annotated and with an Introduction by D'Orsay W. Pearson



RUTHERFORD • MADISON • TEANECK
FAIRLEIGH DICKINSON UNIVERSITY PRESS

LONDON: ASSOCIATED UNIVERSITY PRESSES

Portada de una traducción reciente al inglés (1977) de Reprovación de las supersticiones y hechizerías, *de Ciruelo.*

sunt operari multa mirabilia secundum quid, seu respectu nostri, daemoni adiuvante, et Deo Permittente", afirma Martín del Río con el más puro estilo escolástico. Gran parte de sus pretendidas manifestaciones eran meras supercherías, "la inmensa mayoría de las acciones mágicas y maravillosas de las que se habla —dice Francisco de Vitoria en sus *Relectiones Theologicae* (1557)— son falsas y fingidas, creídas por la vanidad de los hombres". Sin embargo, continúa el mismo autor, "no todas las obras de los magos son falsas y simuladas".

La directa implicación religiosa de la magia negra explica que fuera perseguida con especial dureza por la Inquisición. La desconfianza ante la magia en su conjunto se manifiesta en sucesos como la denuncia en 1583 de su enseñanza en la Universidad de Valladolid, aunque se tratara en principio de la "magia natural", prevista como materia docente en los estatutos. Se creía que el nigromante, en la práctica, "va mezclando lo uno con lo otro, que son la mágica natural con la del demonio", según expresión de Antonio de Torquemada (1570).

Los nigromantes de tipo culto denunciados ante la Inquisición pertenecían a una variada gama comprendida entre dos extremos que podemos personificar en un par de personajes muy conocidos y que Caro Baroja ha estudiado de modo espléndido.

El primero es el médico Eugenio de Torralba, cuya vida serviría de tema a la literatura de creación —desde Cervantes a Campoamor— y a las tradiciones populares. Formado en Roma en un ambiente ocultista, todas sus acciones mágicas las achacaba a un espíritu "familiar" llamado Zaquiel. Durante el largo proceso al que fue sometido por la Inquisición de Cuenca (1528-1531), la clave de su defensa fue demostrar que se trataba de un "ángel bueno" y en modo alguno de un espíritu diabólico, sin pretender negar la índole sobrenatural de los hechos.

El segundo es el clérigo Amador de Velasco, que fue denunciado ante el Santo Oficio en Madrid por uno de sus iniciados (1576). En el polo opuesto a la actividad de Torralba, basó su defensa en convencer al tribunal que se había limitado a aplicar procedimientos meramente *naturales*, expuestos en libros permitidos o "vendibles", y a presentarse como un estudioso movido por la curiosidad.

Velasco tenía en su poder un recetario mágico, cuyo índice es uno de los documentos más notables de la necromancia en la España de esta época. En su biblioteca figuraba también un ejemplar de la *Clavicula Salomonis*, la célebre síntesis de toda una serie de escritos sobre la materia. A pesar de su prohibición expresa en los índices inquisitoriales, la *Clavicula* alcanzó amplia difusión en nuestro país, a través de versiones latinas impresas fuera de España o de copias manuscritas en romance de contenido mucho más corrompido.

En relación con este texto, parece obligada la mención de Lorenzo Ferrer Maldonado como representante de un tipo distinto: el aventurero a caballo entre la subcultura marginada y la ciencia académica. Afirmaba, por una parte, "que él sólo sabía el secreto para descifrar la Clavícula de Salomón, con lo cual venía a alcançar y perfiçionar el verdadero lapis, nunca jamás enteramente hallado de los alchimistas en tantos siglos". Por otra, intentó engañar al Consejo de Indias, afirmando que había descubierto un paso —que llamaba "estrecho de Anian"— que permitía una navegación rápida y segura hasta las Filipinas y las Molucas. Publicó, a pesar de ello, un notable tratado de cosmografía y arte de navegar, completamente de acuerdo con los patrones del saber académico.

Cartografía cromosómica con marcadores de ADN

Diversas secuencias del ADN de los cromosomas humanos actúan como marcadores genéticos. Permiten localizar genes defectuosos y, en conjunto, configuran un mapa cromosómico

Ray White y Jean-Marc Lalouel

Sea el caso de una enfermedad hereditaria que se transmite según las reglas mendelianas clásicas. Entre los 100.000 genes, repartidos en 23 pares de cromosomas, que constituyen la dotación humana debe haber uno defectuoso. Los síntomas y el desarrollo de la enfermedad se han descrito con minucioso detalle, pero se desconoce su base bioquímica e incluso resulta imposible predecir con exactitud quién padecerá la enfermedad. No nos estamos refiriendo a unas pocas enfermedades de escasa incidencia, sino a la mayoría de las 3000 enfermedades congénitas conocidas, incluidos azotes familiares como la enfermedad de Huntington y la fibrosis cística. ¿Por dónde empezar a investigar los mecanismos que causan tales enfermedades? ¿Cómo diseñar un ensayo para diagnosticarlas? Y, en última instancia, ¿cómo tratarlas?

Ahora puede ya iniciarse la batalla asaltando el propio gen defectuoso. El territorio a explorar es amplio: los cromosomas humanos están formados por moléculas lineales de ADN bicatenario de una longitud total aproximada de tres mil millones de pares de bases (subunidades químicas que cifran la información a lo largo del ADN). En comparación, un gen típico, esto es, una unidad completa de información genética, es un segmento muy pequeño, que abarca sólo unos 10.000 pares de bases. A pesar de todo, si se correlaciona la herencia de un segmento concreto de ADN –un "marcador"– con la herencia de una enfermedad, puede localizarse el gen mutante hasta en una distancia de uno o dos millones de pares de bases, esto es, menos de la milésima parte del genoma humano (el complemento total de ADN). Con ese grado de precisión, el gen queda al alcance de las herramientas moleculares, lo que posibilita clonarlo y examinar su actividad. La identificación de un marcador genético estrechamente ligado a una enfermedad permite, además, seguir la herencia del gen que causa dicha enfermedad. Ello abre la vía al desarrollo de un ensayo sencillo de detección de portadores y futuras víctimas de la enfermedad.

La estrategia básica, conocida por análisis de ligamiento, constituye una de las herramientas más antiguas de la genética clásica. En nuestro laboratorio de la Universidad de Utah, y en otros muchos, dicha estrategia ha cobrado nuevo impulso gracias a las posibilidades que ofrecen las nuevas técnicas de la biología molecular, que facilitan el acceso a un amplísima gama de marcadores: los RFLP (por *restriction-fragment length polymorphism*, polimorfismo del tamaño de los fragmentos de restricción). Merced al análisis del ligamiento se conocen ya marcadores RFLP para varios genes causantes de enfermedades; la misma estrategia podrá aplicarse a otras muchas afecciones. Este tipo de aproximación experimental sirve, además, a una causa más general. Mediante el seguimiento simultáneo de la herencia de muchos marcadores RFLP en familias sanas, nuestro grupo y otros hemos empezado a determinar sus posiciones relativas y localizarlas en el marco físico del cromosoma. El objetivo final es trazar un mapa completo de marcadores, esto es, un conjunto ordenado de puntos de referencia que cubra todo el genoma y localice genes patógenos con una eficacia muy superior de lo que permiten los marcadores aislados.

La estrategia del ligamiento se aprovecha del procedimiento por el que se heredan los genes. Una célula humana normal posee 23 pares de cromosomas homólogos. De cada pareja cromosómica, un miembro procede de la madre y el otro del padre. Durante la meiosis, las divisiones celulares que forman las células germinales (espermatozoos u óvulos), los cromosomas homólogos de una célula progenitora se duplican y se distribuyen entre cuatro células germinales, cada una de las cuales sólo recibe 23 cromosomas. Sin embargo, los cromosomas parentales no se transmiten intactos. En el transcurso de la meiosis, los cromosomas homólogos se recombinan, esto es, se "entrecruzan " y permutan segmentos de igual longitud [*véase la figura 3*]. Resulta, así, que los cromosomas que reciben las células germinales suelen estar formados por retazos de los dos cromosomas parentales. Gracias a la recombinación podemos detectar el ligamiento existente entre un marcador y una enfermedad.

Las numerosas diferencias entre los miembros de cada pareja de cromosomas homólogos permiten detectar la recombinación y aprovecharla en el análisis del ligamiento. Es frecuente que porten alelos distintos, o versiones diferentes, de muchos de sus genes, o de las secuencias de ADN aparentemente sin sentido que se encuentran entre los genes y en su interior. Los cromosomas recombinantes que se reparten a las células germinales durante la meiosis representan nuevas combinaciones de tales secuencias. Pueden combinarse y transmitirse juntos alelos que, perteneciendo a loci distintos, se encuentren cada uno en un miembro distinto de la pareja de homólogos. Y pueden separarse dos alelos, también de distintos loci, presentes antes en un mismo cromosoma, de tal suerte que sólo se herede uno de ellos.

Cuanto más próximos estén dos loci en un mismo cromosoma parental, con menos frecuencia se separarán sus alelos al permutarse ADN entre los cromosomas homólogos durante la meio-

sis. Así, para precisar la distancia que media entre un gen de particular interés —por ejemplo, un alelo mutante causante de una enfermedad— y un marcador, se correlaciona el patrón de herencia de sus alelos. Si en una familia afectada los individuos que desarrollan la enfermedad heredan casi siempre la misma versión del marcador, el gen mutante y el marcador se encontrarán en el mismo cromosoma, y muy próximos. Se dice entonces que el marcador y el gen patógeno están ligados.

Los marcadores que estén más alejados del gen se recombinarán con él más a menudo. En tales casos, será menos probable que la enfermedad se herede conjuntamente con algún marcador. En el caso más extremo, cuando un marcador y un gen patógeno, instalados incluso en el mismo cromosoma, están muy alejados, su frecuencia de recombinación se aproximará al 50 por ciento. Se dice entonces que gen y marcador no están ligados: tienen la misma probabilidad de heredarse juntos que separados. El mismo patrón del 50 por ciento de coheredabilidad se registra cuando un marcador y un alelo mutante están situados en cromosomas distintos.

Para relacionar la herencia de un marcador con una enfermedad deben cumplirse dos condiciones: el marcador resultará fácilmente detectable y en la población habrá varias versiones distintas del mismo. Sólo puede detectarse el ligamento si quien lleva los alelos normal y mutante de un gen patógeno porta también dos versiones distintas del marcador. Si los dos alelos del marcador son indistinguibles, no podrán detectarse en la descendencia los entrecruzamientos entre el marcador y el gen de la enfermedad. No habrá forma de distinguir un marcador ligado de otro que no lo esté.

Hasta hace pocos años, sólo unos pocos marcadores cumplían ambos requisitos. Los genes que determinan ciertas enzimas, antígenos de grupos sanguíneos y otras proteínas poseen múltiples alelos, que se manifiestan en el polimorfismo proteico: la presencia de versiones distintas y detectables de la proteína que determina cada gen. Sin embargo, se conocía sólo 25 o 30 sistemas de marcadores con esas características que, además, únicamente cubrían pequeñas secciones de unos pocos cromosomas. Por falta de marcadores, la mayor parte del genoma humano permanecía inaccesible al análisis de ligamiento.

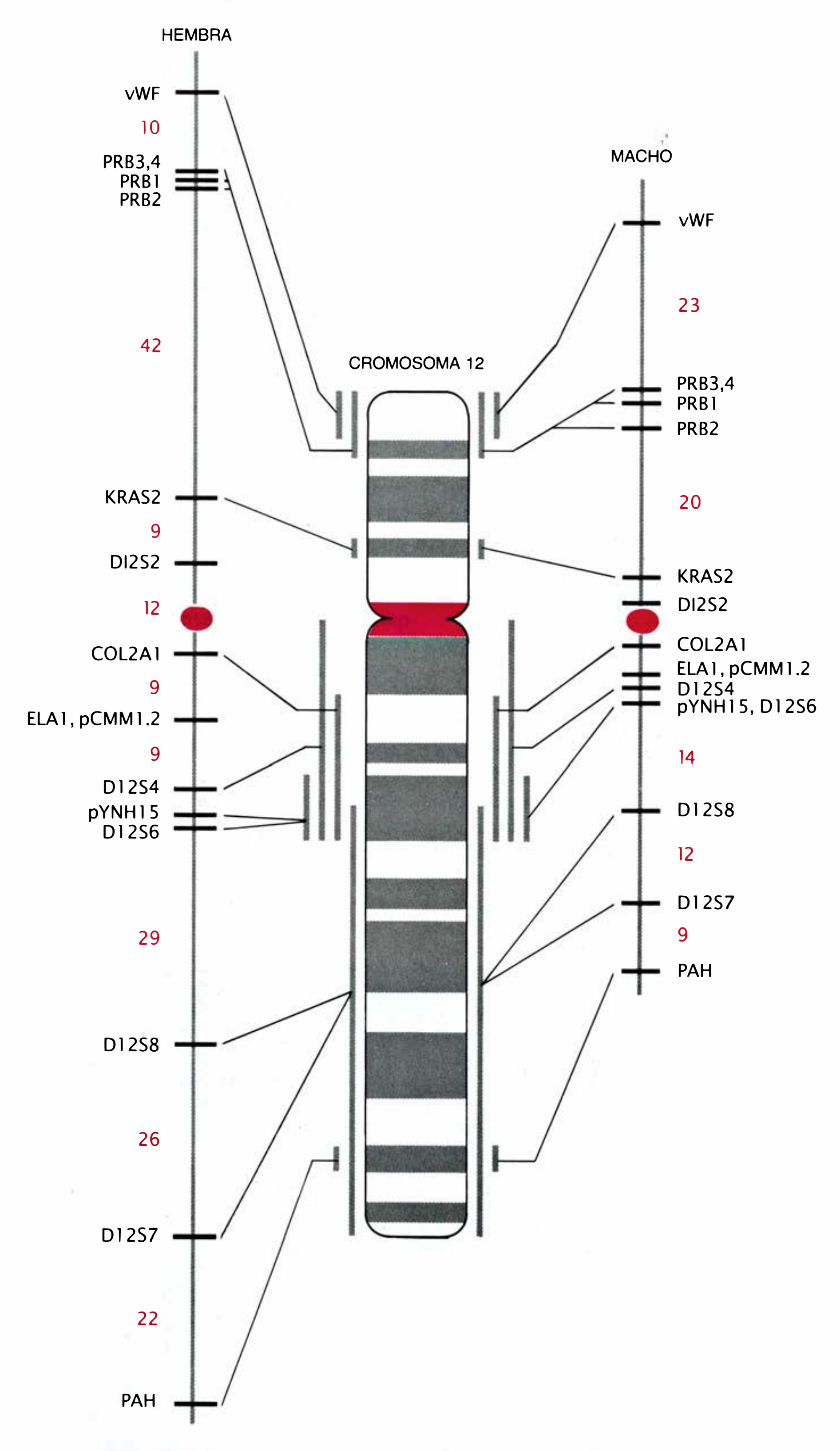
Con la llegada de las técnicas de recombinación de ADN *in vitro*, a mediados de la década de 1970, la cartografía basada en los estudios de ligamiento se transformó, para beneficio de la genética humana, en una herramienta poderosa. Los orígenes de dicha transformación pueden situarse en abril de 1978, con ocasión de un seminario de genética patrocinado por la Universidad de Utah. David Botstein, del Instituto de Tecnología de Massachusetts, Ronald W. Davis, de la Universidad de Stanford, y Marc H. Skolnick, de Utah, lanzaron la idea de que la propia secuencia de ADN podía ser fuente de numerosos marcadores fácilmente detectables. Uno de nosotros (White) captó enseguida las enormes posibilidades de esa nueva aproximación y decidió someter a prueba la hipótesis, comprometiendo a su laboratorio en el desarrollo de una serie de marcadores de ADN que permitiesen detectar ligamiento en cualquier lugar del genoma humano. Botstein, White, Skolnick y Davis publicaron en 1980 el primer artículo que pormenorizaba el desarrollo de esa aproximación experimental. Mientras tanto, otros muchos investigadores empezaron a encontrar marcadores en el ADN humano y a especular sobre su utilidad. La idea se había convertido ya en realidad.

El poder resolutivo de la nueva estrategia de ligamiento radica en el altísimo nivel de polimorfismo normal que contiene la secuencia de pares de bases que forman el ADN. Entre las secuencias de dos cromosomas homólogos se observa, en promedio, una diferencia cada 200 a 500 pares de bases. La detección de esas variantes alélicas podría constituir una fuente virtualmente ilimitada de marcadores, repartidos por todos los cromosomas humanos.

Las herramientas moleculares que responden al nombre de enzimas de



1. FAMILIAS NUMEROSAS con abuelos vivos (versiones modernas de esta de principios de siglo) constituyen el material ideal para los estudios de ligamiento genético. Se deduce la posición que ocupan determinados puntos cromosómicos a partir de la frecuencia con que las variaciones que afectan a dichos sitios se transmiten juntas, de padres a hijos. El estudio de las pautas de herencia de una enfermedad genética y de marcadores genéticos arbitrarios permite asignar al gen de la enfermedad su lugar en el cromosoma. Mediante la correlación de la herencia de muchos marcadores se dibujan los mapas cromosómicos.



2. MAPA DEL CROMOSOMA 12, elaborado siguiendo las pautas hereditarias de marcadores de ADN: sitios en los que las dos copias de un cromosoma suelen llevar secuencias de ADN comprobablemente diferentes. La ordenación de los marcadores que se acepta es la que presenta mayor probabilidad estadística. Los marcadores están separados por distancias que reflejan su frecuencia de recombinación, es decir, el porcentaje de veces que las versiones de los marcadores presentes en el mismo cromosoma parental se separan merced a un proceso de recombinación genética durante la formación de óvulos o espermatozoos. La frecuencia de recombinación entre dos marcadores aumenta con la distancia física que les separa, si bien la relación exacta entre frecuencia de recombinación y distancia varía en función de varios factores, entre ellos el sexo. En el cromosoma 12, por ejemplo, la tasa total de recombinación es mayor cuando el cromosoma lo transmite una mujer. Esa es la razón de que, en este caso, el mapa genético de la mujer sea mayor que el del hombre. En algunos marcadores se ha determinado su localización cromosómica aproximada (*centro*).

restricción, o restrictasas, facilitan dicha detección. Cada restrictasa, sintetizada por una especie bacteriana, se une al ADN cuando encuentra una secuencia de pares de bases breve y específica; corta la molécula por un sitio de la secuencia también específico. Cualquier variación de la secuencia de ADN que cree o elimine un sitio de restricción alterará la longitud del fragmento o fragmentos de ADN resultantes. La variación origina un polimorfismo del tamaño de los fragmentos de restricción, esto es, un RFLP.

El RFLP define un marcador en potencia. Una restrictasa, sin embargo, puede detectar millones de sitios de corte por todo el ADN humano. ¿Cómo discernir uno o dos fragmentos entre los millones posibles? En primer lugar, los fragmentos se separan mediante electroforesis: un campo eléctrico los distribuye por un gel, resultando su movilidad en él inversamente proporcional a la longitud que tengan. Una técnica muy sensible y de gran poder de resolución, denominada transferencia de Southern ("Southern blotting", en honor de Edward M. Southern, que la desarrolló en la Universidad de Edimburgo), identifica los fragmentos que interesen.

La técnica de Southern se apoya en el carácter singular de la molécula de ADN. A lo largo de las dos cadenas del ADN, las bases se aparean de acuerdo con ciertas reglas, de tal suerte que la secuencia de una cadena constituye la única posibilidad de apareamiento para la secuencia de la otra. En una muestra de ADN "desnaturalizado", esto es, calentado o expuesto a un *p*H alto para separar sus cadenas, un fragmento de ADN de cadena sencilla puede actuar de sonda que busque y se una a su secuencia complementaria. En la técnica de Southern, una vez separados los fragmentos de ADN por electroforesis, se desnaturalizan y transfieren del gel a una membrana, donde se ponen en contacto con una sonda de ADN marcada con un isótopo radiactivo. La sonda sólo se hibrida, se une, al fragmento o fragmentos que complementan su secuencia de bases. El marcaje radiactivo permite detectar la posición de los fragmentos y, por tanto, sus tamaños.

Para detectar un RFLP es preciso, pues, encontrar una sonda que sea complementaria a una zona de ADN cercana al sitio de corte de la restrictasa. Normalmente se suele elegir, al azar, un segmento de ADN de una colección ("genoteca") de fragmentos de ADN clonados, representativos de la

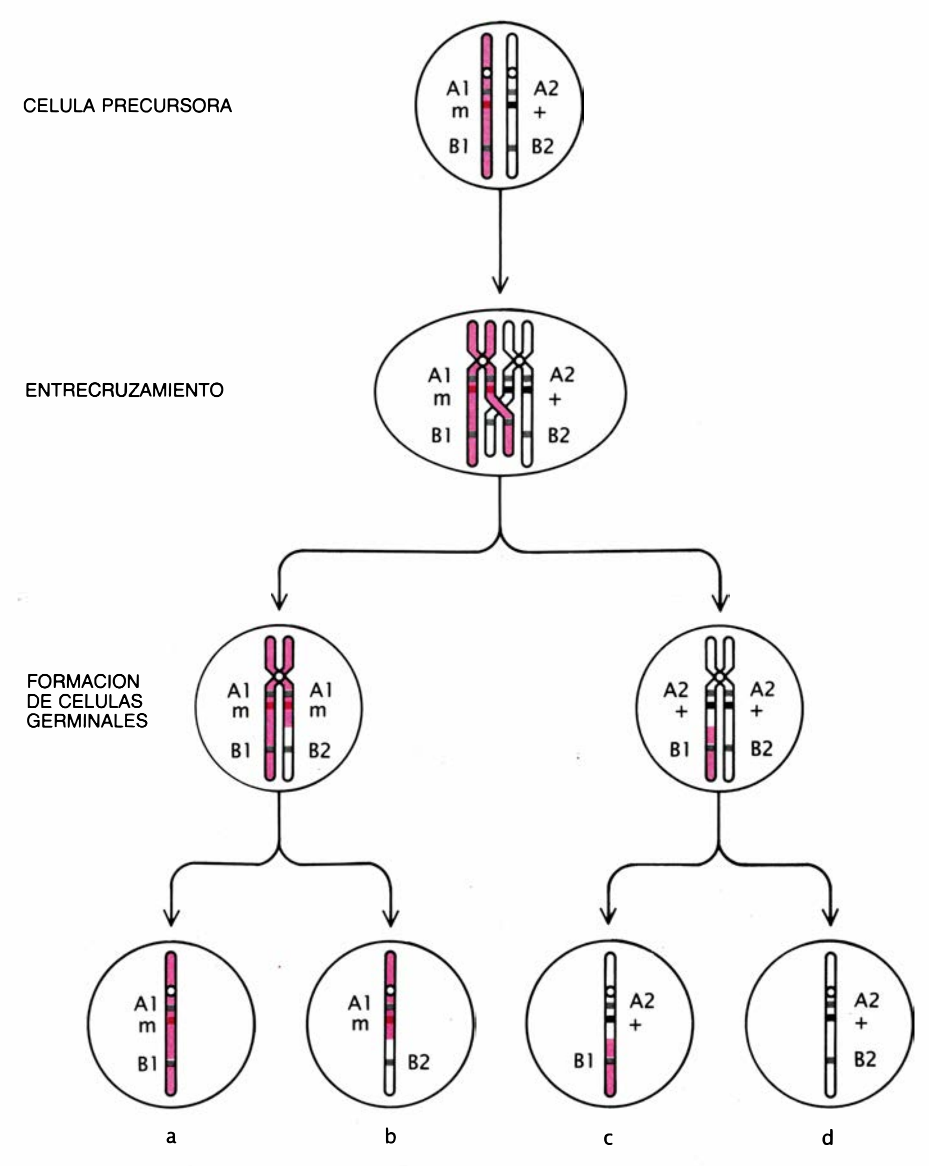
totalidad del genoma humano. Se desnaturaliza, se marca radiactivamente y se pone en contacto con la membrana que contiene los fragmentos de ADN, tranferidos tras la digestión con la restrictasa y posterior separación en un gel. Si se realiza el experimento con ADN de varios individuos distintos, y las bandas radiactivas aparecen en lugares diferentes, el ADN clonado ha detectado la variabilidad de cortes típica de un caso de polimorfismo de ADN. La sonda y el RFLP que dicha sonda detecta constituyen un marcador genético, que nos brinda un punto de referencia en el genoma: un breve segmento de ADN polimórfico cuya pauta de herencia puede seguirse fácilmente.

Este marcador de ADN, definido por el RFLP, se encuentra, en una u otra forma, en todos los individuos, padezcan o no la enfermedad. Pero si en una familia una enfermedad genética se transmite siempre junto con cierto alelo del RFLP, puede postularse que el gen mutante se encuentra en la misma región cromosómica que el marcador. En otra familia afectada, el mismo marcador mostrará tambien ligamiento, si bien la forma concreta del marcador que acompaña a la enfermedad puede ser distinta. Por supuesto, el ligamiento con un marcador de ADN arbitrario no informa sobre la posición física del gen, aunque en muchos casos (por ejemplo, para las pruebas de diagnóstico) su localización física es indiferente. Con todo, la sonda permite también identificar el cromosoma que porta el marcador y el gen de la enfermedad. Si la sonda se mezcla con un juego completo de cromosomas humanos, por ejemplo, se hibridará con el cromosoma donde se halle el sitio marcador.

El valor de cualquier marcador depende en gran medida de cuántas variantes del mismo posea la población. Cuantas más versiones existan, más probable será que un individuo que lleve un gen causante de alguna enfermedad presente dos alelos distintos en el locus de interés. Cuando eso ocurre, puede detectarse la recombinación entre el marcador y la enfermedad en los descendientes. Muchos RFLP tienen su origen en el cambio de una sola base, o en la adición o deleción de unos pocos pares de bases en el sitio por donde corta la restrictasa. Este tipo de variación ejerce un efecto sencillo: el sitio de restricción aparece o está ausente. El RFLP, en este caso, presenta sólo dos formas, de suerte que al menos la mitad de la población será homocigótica para el locus marcador, es decir, tendrá la misma variante en ambos cromosomas homólogos. (En ocasiones, dos sitios de restricción están tan cerca que pueden detectarse con la misma sonda. Tales casos rinden un marcador con cuatro alelos.)

Existe otro tipo de polimorfismo de ADN que genera muchas versiones distintas de un RFLP. Repartidas por el ADN humano se encuentran secuencias que no determinan proteína alguna y que se repiten muchas veces. Se ignora el origen y significado de esas "repeticiones en tándem". Pero las técnicas de cartografía basadas en el ligamiento también pueden servirse de ellas, pues el número de veces que se repiten en un locus determinado varía de unas cuantas copias a centenares de ellas. En correspondencia, se modificará también la longitud de los fragmentos de restricción que se generan al cortar por un sitio cercano a esas repeticiones en tándem [*véase la figura 5*]. En tales casos, los RFLP presentarán muchas formas distintas, en vez de sólo dos. Dada esa variabilidad del conjunto de la población, resulta bastante probable que un mismo individuo porte versiones diferentes del RFLP en cromosomas homólogos. Si es así, con la técnica de Southern se observarán dos fragmentos de distintas logitudes, procedentes de los dos cromosomas homólogos.

Se obtienen de un modo más sistemático que las sondas para marcadores



3. RECOMBINACION, que permite detectar el ligamiento genético. Se representa el paso de una pareja de cromosomas homólogos a través de la meiosis, proceso que da lugar a las células germinales (óvulos y espermatozoos). Los cromosomas portan alelos diferentes para dos marcadores (*A*, *B*). Uno de los cromosomas lleva, además, un alelo mutante de cierto gen patógeno (*m*); el cromosoma homólogo lleva el alelo normal (+). En la célula precursora, la enfermedad está asociada a los alelos *1* de ambos marcadores, *A* y *B*. Durante la primera fase de la meiosis, los cromosomas se replican. Los cromosomas homólogos se "entrecruzan", intercambiando segmentos de igual longitud. Aquí, el entrecruzamiento tiene lugar entre los loci *A* y *B*. Resultan de ello dos células germinales (*a*, *d*) con las combinaciones alélicas parentales y dos (*b*, *c*) que portan cromosomas recombinantes. En la célula *b*, el gen mutante va acompañado aún del alelo *1* del locus *A*, pero ahora en el locus *B* se encuentra el alelo *2*. Una frecuencia baja de entrecruzamientos entre aquel gen y el locus *A* indicaría que la enfermedad y el marcador genético se encuentran estrechamente ligados.

GEN PATOGENICO

MARCADOR

A B C D

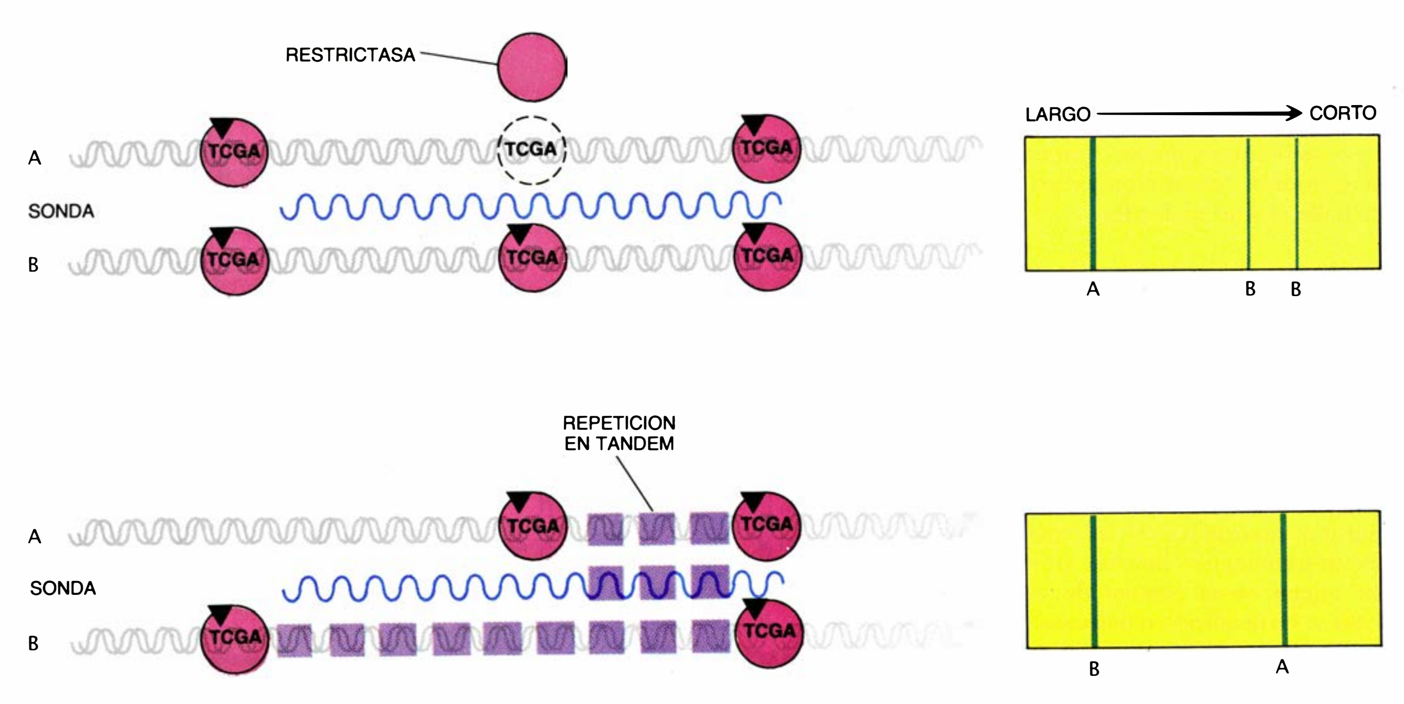
MEIOSIS

A D A C B D B C A/B C

NO RECOMBINANTES

RECOMBINANTES

4. LIGAMIENTO entre el gen patógeno y un marcador, evidenciado al estudiar la historia familiar de la enfermedad. Se muestran las características genéticas de una pareja hipotética y de sus hijos. Uno de los padres sufre una enfermedad genética causada por un único alelo mutante (*rojo*). El otro es sano y, por tanto, lleva la versión normal del gen (*rosa*). El hijo que hereda la enfermedad suele también heredar cierto alelo marcador (*púrpura*) del progenitor enfermo.



5. MARCADORES DE ADN, sitios en los cuales suelen diferir en su secuencia de ADN los cromosomas homólogos. Se detectan como RFLP (polimorfismo del tamaño de los fragmentos de restricción). El ADN se digiere con una restrictasa, enzima que corta donde encuentra una breve secuencia específica de nucleótidos (en este caso *TCGA*). Un posible tipo de marcador (*arriba, a la izquierda*) es aquel en el que una diferencia de la secuencia determina la presencia de un sitio de restricción en uno de los cromosomas y su ausencia en el homólogo. Los fragmentos de restricción que resulten de cada cromosoma diferirán en longitud. Una sonda de ADN cuya secuencia de bases sea complementaria a la del marcador revelará la presencia de los fragmentos, una vez separados por electroforesis (*arriba, a la derecha*). Otro marcador es el caracterizado por una VNTR, esto es, una variación del número de repeticiones en tándem (secuencias de ADN cortas y muy repetidas). La distancia entre los puntos de corte varía de un cromosoma homólogo a otro, originando fragmentos distintos (*abajo, a la derecha*).

normales aquellas otras que se basan en las variaciones del número de repeticiones en tándem (VNTR). Alec J. Jeffreys, de la Universidad de Leicester, ha reconocido recientemente que las secuencias repetidas de muchos loci VNTR, de partes distintas del genoma, presentan similitudes. De nuevo, la explicación evolutiva resulta poco clara, pero la homología parcial entre secuencias conlleva la posibilidad de que, bajo ciertas condiciones, pueda utilizarse una sonda complementaria a un locus VNTR para seleccionar sondas específicas de otros loci de una genoteca de ADN clonado. De los casi 600 marcadores de ADN desarrollados hasta ahora en nuestro laboratorio, unos 300 son VNTR.

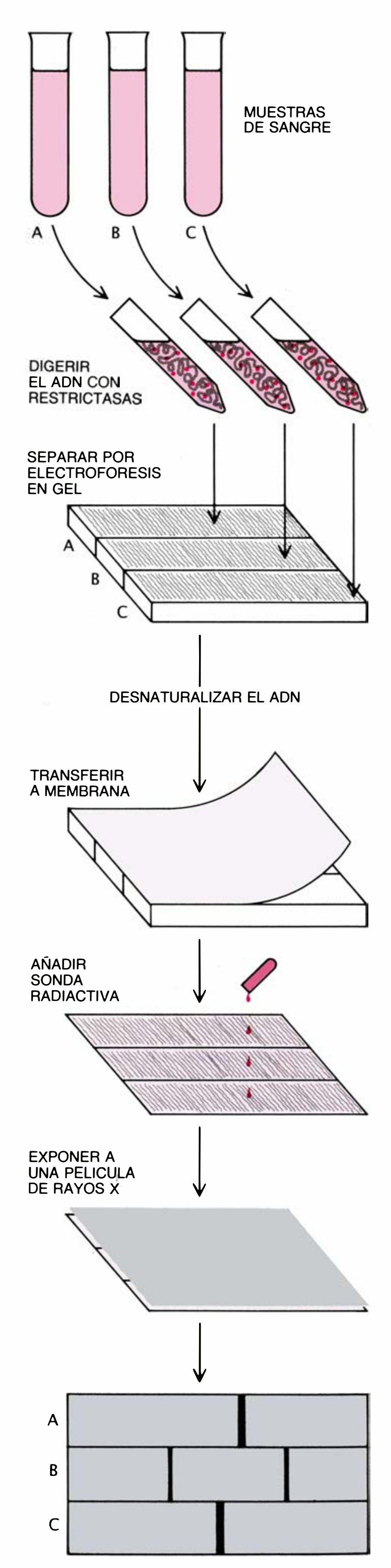
Esos marcadores pueden utilizarse como elementos de referencia en un mapa general del genoma, o servir para un propósito inmediato, como seguirle la pista al gen responsable de tal o cual enfermedad. Descubrir un marcador cuya herencia se correlacione con la aparición de una enfermedad puede convertirse en un empeño difícil cuando se trabaja con cromosomas no cartografiados. Lo usual es empezar la investigación ignorando la localización cromosómica del gen de la enfermedad, o de cualquier marcador cuya pauta de herencia se esté siguiendo en una familia afectada; por ello, cabe la posibilidad de rastrear el ligamiento de decenas de marcadores muy alejados del gen que interesa y pasar por alto otros que sí están ligados. No obstante, la estrategia basada en el ligamiento ha cosechado ya éxitos notables.

Si se conoce qué cromosoma investigar, el número de marcadores a reseguir cae de unos centenares a media docena, aproximadamente. Pensemos, por ejemplo, en una enfermedad congénita que casi siempre afecte a los varones pero se herede a través de la madre: es muy probable que la cause un gen recesivo del cromosoma *X*, determinante del sexo. (Las madres portadoras disponen de otro cromosoma *X* dotado de una copia normal del gen, que enmascara al gen recesivo causante de la enfermedad; si la mutación la hereda un hijo, que sólo porta ese cromosoma *X*, desarrollará los síntomas de la enfermedad.) Para localizar el gen, basta con ensayar los marcadores cuya presencia en el cromosoma *X* se haya demostrado.

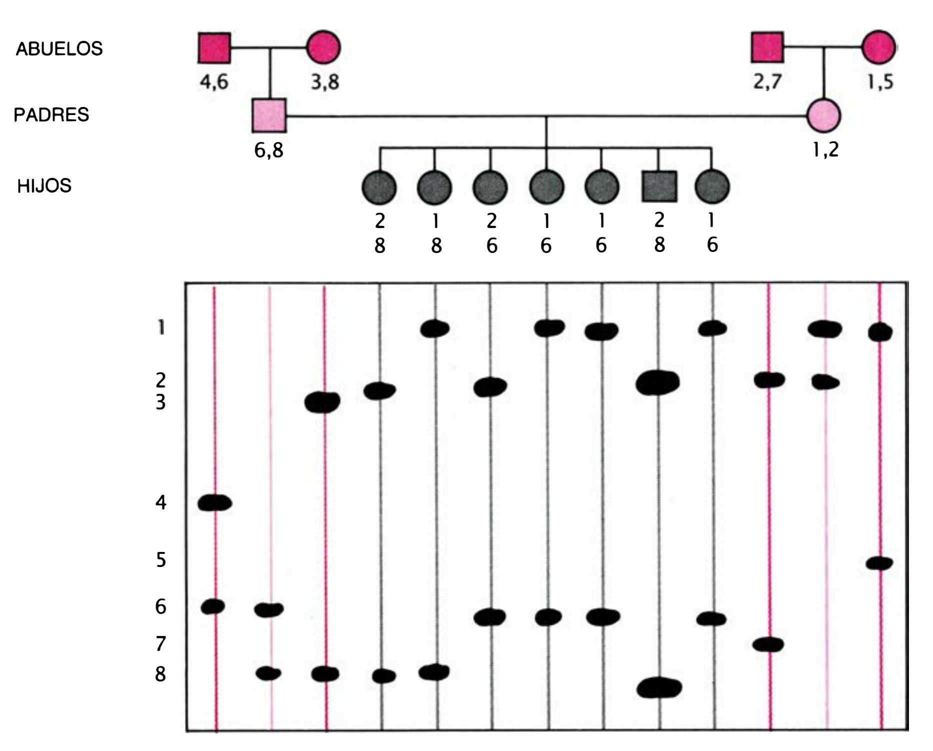
Los genes de enfermedades ligadas al cromosoma *X* figuran entre los primeros cuyo seguimiento se realizó con la técnica de los RFLP. Se empezó por el gen que causa la distrofia muscular de Duchenne, y probablemente también la distrofia muscular de Becker (cartografiado por Kay Davies, de la Universidad de Oxford, y Robert Williamson, del Hospital St. Mary de Londres). También un número cada vez mayor de enfermedades causadas por defectos localizados en cromosomas autosómicos (los 22 pares de cromosomas no sexuales) se han sometido a la estrategia del ligamiento.

La enfermedad de Huntington fue la primera de carácter autosómico que se relacionó con un marcador de ADN. El mérito fue de James F. Gusella y sus colegas, de la Universidad de Harvard, que estudiaron varias familias afectadas por ese mal, en Estados Unidos y cerca del lago Maracaibo, Venezuela. El equipo tuvo mucha suerte, pues bastó el ensayo de ocho marcadores para dar con uno ligado a la enfermedad. Desde entonces, y gracias a los trabajos realizados por varios laboratorios, entre ellos el nuestro, se han descubierto marcadores para genes causantes de distintas enfermedades: fibrosis cística, neurofibromatosis periférica, o enfermedad de Von Recklinghausen (trastorno caracterizado por la presencia de manchas cutáneas de color "café con leche" y por una tendencia a desarrollar tumores y otras alteraciones óseas y del sistema nervioso), y poliposis cólica hereditaria (cuyas víctimas desarrollan múltiples pólipos en el colon y poseen un alto riesgo de cáncer de colon). Y, lo que puede ser más intrigante, se ha descrito ligamiento hereditario para ciertas formas de la enfermedad de Alzheimer y de síndromes maniaco-depresivos que cursan en familias.

Un "golpe de suerte" puede abrir el camino para la identificación del gen, primer paso para investigar el mecanismo molecular de la enfermedad. Si se clona el gen y se determina su secuencia de bases es posible deducir la composición de la proteína que determina,



6. ANALISIS DEL RFLP a partir de una muestra de sangre. Se extrae el ADN de los núcleos de los leucocitos y se digiere con una restrictasa. Los fragmentos que resultan se separan mediante electroforesis en gel. El RFLP se detecta aplicando la técnica de transferencia de Southern. Primero se calienta el ADN en el gel, para desnaturalizarlo, esto es, para que se separen sus cadenas. A continuación, se transfiere a una membrana de nylon, que se incuba con una sonda: un segmento de ADN monocatenario complementario al locus del RFLP marcado radiactivamente. La sonda se hibrida con los fragmentos del locus. Una película de rayos X colocada sobre la membrana detecta los fragmentos radiactivos y revela qué versión del RFLP se halla presente. En los análisis familiares se suelen estudiar al mismo tiempo muestras de ADN tomadas de varios individuos.



7. COMPORTAMIENTO HEREDITARIO DE UN RFLP. Se deduce comparando los fragmentos de restricción de varios miembros de la familia. El marcador RFLP analizado en estas tres generaciones familiares (un locus VNTR) posee muchos alelos, definidos por un fragmento de restricición de tamaño específico. Todos los individuos que recoge este árbol genealógico (donde los cuadrados representan a los machos y los círculos a las hembras) portan dos alelos distintos del marcador, uno en cada cromosoma homólogo. Los niños heredan un alelo de cada progenitor. Si en una familia la presencia de la enfermedad se correlaciona con la presencia de uno de los alelos del RFLP, resulta muy probable que ambos se encuentren ligados.

y quizá también identificar un defecto específico. Una vez establecida la secuencia, se puede sintetizar la proteína y fabricar anticuerpos contra ella en animales de laboratorio. Marcados adecuadamente, esos anticuerpos pueden utilizarse para conocer la distribución de la proteína en los tejidos afectados por la enfermedad, información que puede resultar decisiva para diseñar un tratamiento de la enfermedad.

En muchos casos, sin embargo, la imprecisa localización inicial impide acercarse al gen con las técnicas moleculares disponibles. Por ejemplo, el gen responsable de la enfermedad de Huntington se recombina con el primero de sus marcadores identificados con una frecuencia de aproximadamente el 5 por ciento. Ello quiere decir que el marcador puede encontrarse a unos cinco millones de pares de bases de distancia del gen. Para identificar y clonar un gen, la distancia debe reducirse a un millón de pares de bases, aproximadamente, esto es, hay que contar con marcadores que se recombinen con el gen con una frecuencia del 1 por ciento. Lo ideal es poseer marcadores a ambos lados del gen, que cerquen el segmento de ADN a analizar.

Se dispone ya de marcadores estrechamente ligados que flanquean los genes causantes de la fibrosis cística, neurofibromatosis periférica y poliposis hereditaria. También se ha identificado un nuevo marcador estrechamente ligado al gen de la enfermedad de Huntington. La búsqueda de los genes responsables de esas enfermedades va por buen camino. Aunque el tipo de aproximación experimental varía, una táctica común consiste en "peinar" una genoteca de fragmentos cromosómicos clonados, a la búsqueda de alguno que reconozcan las sondas para los dos marcadores periféricos. Una vez conseguido el fragmento, que presumiblemente contendrá los dos marcadores y el gen situado entre ambos, se divide. Se clonan los trozos y se analiza la actividad biológica de todos ellos. Generalmente, el fragmento se utiliza como sonda de ARN mensajeros (cuya presencia indica que el gen se está expresando) de los tejidos afectados por la enfermedad. Si con el fragmento se detecta un ARN mensajero exclusivo del tejido interesado, la propia sonda puede contener parte o todo el gen responsable de la enfermedad.

Mediante una estrategia distinta se ha identificado el defecto genético responsable de la distrofia muscular de Duchenne. La región del cromosoma *X* que Davies y Williamson habían asociado con la enfermedad contiene segmentos que faltan en muchos pacientes. La conclusión inmediata es que la enfermedad puede derivar de la ausencia de un gen normal, o de parte de él. Tras identificar la región común ausente en las víctimas de la enfermedad, Louis M. Kunkel, de la Facultad de Medicina de Harvard, y sus colegas aislaron y clonaron el gen.

A veces, antes incluso de identificar el gen de una enfermedad, el propio ligamiento sugiere posibles mecanismos causales. Quizás el marcador se encuentre junto a un gen de función conocida, del que se sospecharía entonces alguna implicación en la enfermedad. Así, por ejemplo, el marcador de la neurofibromatosis periférica está situado en el cromosoma 17, donde se encuentra también el gen que determina el receptor celular del factor de crecimiento neuronal (una sustancia imprescindible para la supervivencia y crecimiento de las células nerviosas). Se consideró, a ese gen, posible responsable de la neurofibromatosis, si bien luego se comprobaría que le separaba cierta distancia del locus de la enfermedad. Habrá, pues, que pensar en la implicación de otros genes del cromosoma 17 en el trastorno.

El ligamiento estrecho entre un marcador y una enfermedad permite diseñar pruebas de detección de portadores y posibles víctimas no nacidas. Dada la frecuencia y carácter insidioso de muchas enfermedades genéticas, urge la puesta a punto de esos ensayos. Por ejemplo, en poblaciones de origen norteuropeo, uno de cada 20 individuos es portador del gen de la fibrosis cística. Por tratarse de un gen recesivo, los portadores no manifiestan los síntomas; pero, si se casan dos portadores, sus hijos tendrán un 25 por ciento de probabilidad de heredar los dos genes defectuosos y contraer la enfermedad. La enfermedad de Huntington está determinada por un gen dominante (que se manifiesta aunque el otro gen sea normal), pero sus síntomas no suelen aparecer hasta la edad adulta, cuando la futura víctima, sin saberlo, ha transmitido ya la enfermedad a la mitad de sus hijos.

Para diagnosticar la presencia de un gen patógeno en un posible portador, es preciso analizar antes el ADN de otros miembros de la familia, sanos o enfermos. El objeto de dicho análisis es determinar qué alelo marcador (o alelos, en el caso de una enfermedad recesiva) se hereda con la enfermedad en esa familia. El descubrimiento de un alelo indicador en el ADN de un potencial progenitor revela que corre el riesgo de transmitir la enfermedad. La obtención de muestras de ADN del

feto poco tiempo después de su concepción posibilita el diagnóstico prenatal de la enfermedad, lo cual permite a los padres tomar una decisión fundamentada sobre la continuidad del embarazo. Importa señalar que, en las familias con alto riesgo para alguna enfermedad genética, el diagnóstico fetal está, en realidad, incrementando el número de nacimientos, pues muchas parejas renunciarían a tener hijos si dudaran de que el niño está sano antes de llegar el embarazo a término.

La búsqueda de casos específicos de ligamiento para ciertas enfermedades ha ido en paralelo a la construcción de mapas de ligamiento con marcadores arbitrarios y genes conocidos, ordenados a lo largo de los cromosomas. Gracias a los mapas de ligamiento, la búsqueda de nuevos genes mutantes resulta un trabajo más meditado y sistemático. Disponiendo de un mapa de ligamiento bien documentado, los investigadores empeñados en la localización de un gen causante de una enfermedad pueden elegir y ensayar marcadores que cubran la mayor parte de cada cromosoma. Una vez descubierto un caso de ligamiento que restrinja la localización del gen a un segmento cromosómico específico, pueden ensayarse nuevos marcadores situados en esa región menor hasta dar con un marcador estrechamente ligado que permita posteriores estudios moleculares.

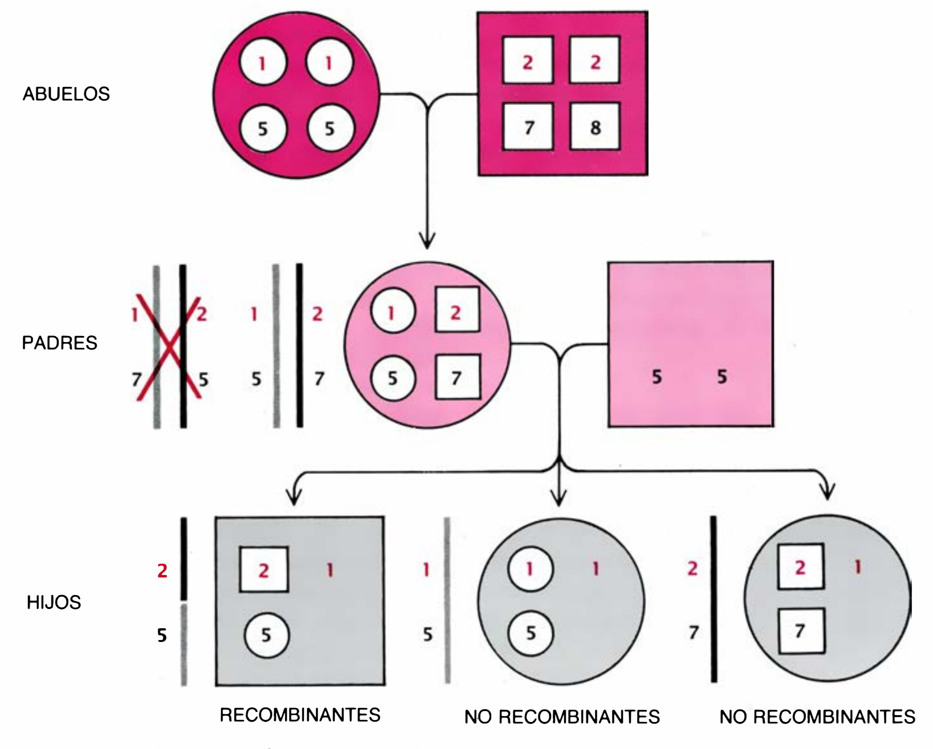
La exploración del genoma en busca de casos de ligamiento no sólo permitirá localizar con mayor precisión los defectos que tienen su origen en la alteración de un solo gen, sino que acelerará también la búsqueda de las bases genéticas de enfermedades de origen plurigénico. Gracias a los mapas de ligamiento se rastrearán muchos sitios cromosómicos simultáneamente, al objeto de hallar alguna pauta hereditaria que se correlacione con la historia familiar de la diabetes, cardiopatías y ciertos tipos de cáncer, enfermedades en las que parece existir sesgo hereditario. Este tipo de enfoque permitiría aproximarse a los genes que confieren predisposición a tales enfermedades.

Con estos mapas se amplían, además, las estrategias de ligamiento: en vez de buscarlo entre un marcador de ADN y una enfermedad, se indaga entre marcadores arbitrarios de ADN. Si los alelos de varios marcadores tienden a transmitirse juntos, deben encontrarse en el mismo cromosoma. La frecuencia con que dichos marcadores se recombinan constituye un reflejo de su "distancia genética".

Los fundamentos teóricos de la cartografía basada en el ligamiento son sencillos, pero esa técnica plantea un difícil reto analítico y de tratamiento de datos. Un mapa de ligamiento a gran escala del genoma, suficiente para localizar con un margen de 10 a 20 millones de pares de bases cualquier gen implicado en una enfermedad, debe incluir entre 100 y 200 marcadores regularmente repartidos. Para disponer de marcadores en intervalos regulares, el mapa necesita un conjunto mucho mayor de marcadores aleatorios. Ello obliga a recoger ADN de centenares de individuos en muchas familias numerosas y comprobar la presencia de RFLP característicos de cada marcador.

El análisis de esa ingente cantidad de datos se complica aún más por el hecho de que quizá dos terceras partes de los marcadores de cualquier individuo no aportan información alguna. Son los casos en los que los dos alelos del marcador en cuestión son idénticos, con lo que no puede detectarse en la descendencia el ligamiento entre dicho marcador y cualquier otro locus. Suele ser frecuente, además, que no se pueda determinar en qué "fase" se encuentran dos marcadores supuestamente ligados, esto es, en qué cromosoma va cada uno. Si se desconoce qué alelos están en el mismo cromosoma del progenitor, no puede establecerse la existencia de recombinación entre los marcadores del hijo.

Estas limitaciones se alivian cuando los datos se recogen entre familias numerosas y formadas con distintas generaciones. En este sentido, nos consideramos muy afortunados, pues para la elaboración de nuestros mapas hemos contado con unos recursos familiares muy importantes. Más de 50 familias del estado de Utah, cada una de ellas de ocho hijos o más, se ofrecieron voluntarias para dar muestras de sangre. A partir de tales muestras aislamos ADN para analizarlo y establecimos líneas celulares permanentes. Interesa en especial que haya muchos niños, pues así se sigue la pista de los cromosomas paternos a través de muchas meiosis, y se calcula la frecuencia de recombinación con mayor fiabilidad que cuando son pocos los hijos. Además, en casi todas las familias de Utah que analizamos vivían los cuatro abuelos. El ADN de los abuelos indica a veces la fase de los marcadores en los padres. Si se sabe, por ejemplo, que uno de los abuelos porta el alelo 1 del marcador *A* y el alelo 3 del marcador *B*, su hijo (uno de los padres) debe portar ambos alelos en el mismo cromosoma si los marcadores están ligados.



8. CON DATOS DE TRES GENERACIONES pueden resolverse posibles dudas sobre las "fases" de la cartografía genética, planteadas por la presencia de dos marcadores en un mismo cromosoma. Si no se conoce en qué fase se encuentran dos marcadores (*color y negro*) en uno de los padres —es decir, de qué modo se distribuyen sus alelos (*números*) entre los cromosomas homólogos— no puede detectarse sin ambigüedad la existencia de recombinación en los hijos. El análisis del ADN de los abuelos (los padres de la madre, en este caso) revelará qué pareja de alelos ha transmitido cada uno de ellos. Al haber recibido la madre los alelos *1* y *5* en el cromosoma heredado de su madre, los alelos *2* y *7* deben encontrarse en el cromosoma que procede de su padre. Por consiguiente, puede descartarse cualquier otra configuración alélica. El primero de los hijos revela, sin ambigüedades, que en el cromosoma de origen materno se ha producido recombinación.

ENFERMEDAD	CROMOSOMA	FECHA
ENFERMEDAD DE HUNTINGTON	4	1983
DISTROFIA MUSCULAR DE DUCHENNE	*X* (GEN CLONADO)	1983
POLICISTOSIS RENAL	16	1985
FIBROSIS CISTICA	7	1985
GRANULOMATOSIS CRONICA	*X* (GEN CLONADO)	1985
NEUROFIBROMATOSIS PERIFERICA	17	1987
NEUROFIBROMATOSIS CENTRAL	22	1987
POLIPOSIS COLICA HEREDITARIA	5	1987
NEOPLASIA ENDOCRINA MULTIPLE IIa	10	1987

9. TABLA DE ENFERMEDADES GENETICAS, donde se incluye una pequeña muestra de aquellas en las que, mediante estudios de ligamiento, se ha localizado el gen defectuoso que las origina. La tabla indica igualmente el cromosoma donde se encuentran el gen y el marcador, así como el año en que se publicó dicho caso de ligamiento. Un ligamiento suficientemente intenso hace del marcador un instrumento de gran utilidad para el diagnóstico de la enfermedad genética entre los miembros de las familias afectadas.

Con todo, las inevitables limitaciones que imponen los datos obligan a construir los mapas sobre bases probabilísticas. Ante las pautas de herencia observadas, puede estimarse con técnicas estadísticas la frecuencia de recombinación más probable entre dos marcadores o, lo que es lo mismo, su distancia genética. Una frecuencia de recombinación del 50 por ciento sugiere que dos marcadores no están ligados. Una frecuencia menor, por ejemplo del 10 por ciento, que presenta un sólido soporte estadístico, indica ligamiento. Pronto, uno de los autores (Lalouel, a la sazón en la Universidad de París) advirtió que nuestra aventura cartográfica exigía disponer de metodología estadística especializada y un tratamiento informático adecuado. Junto con Mark Lathrop, redactó programas especialmente diseñados para almacenar la ingente cantidad de datos y para llevar a cabo el análisis conjunto de las pautas de herencia de muchos marcadores.

Identificada una serie de marcadores ligados, hay que determinar su ordenación en el cromosoma. En principio, podría calcularse, para todas las ordenaciones, la respectiva probabilidad de que genere la pauta hereditaria observada y se eligiría la más verosímil. Ahora bien, con sólo 15 marcadores se pueden obtener hasta $6,5 \times 10^{11}$ ordenaciones diferentes, cifra inmanejable. En la práctica, considerando los marcadores en pequeños grupos, por ejemplo de tres en tres, se eliminan de un golpe familias enteras de ordenaciones improbables.

Por el mero placer de razonar, supóngase que en una familia numerosa se observa la presencia de un grupo de alelos de tres marcadores ligados, A, B y C, que suelen transmitirse agrupados, esto es, o se heredan todos o no se hereda ninguno. Pero, en uno de los hijos, los alelos originales de A y C se heredan con un alelo distinto de B. En otro de los hijos, el alelo original de B se hereda junto con otros alelos distintos de A y C. La ordenación $A-B-C$ es la menos probable, pues en ambos casos necesitaría de una doble recombinación, entre A y B y entre B y C. (En las otras dos ordenaciones alternativas, $A-C-B$ o $B-A-C$, un solo proceso de recombinación explicaría las combinaciones alélicas observadas en los dos hijos.)

Ese es el tipo de estrategia utilizada en los sistemas informáticos que hemos diseñado para, dado un número arbitrario de marcadores ligados, encontrar la ordenación más probable. Una vez establecido el orden más probable para un grupo de marcadores ligados, éstos se asignan a cromosomas. Para ello se hibrida una sonda correspondiente a uno de los marcadores con una serie de cromosomas intactos. Los cúmulos o agregados en ligamiento se enlazan y se va construyendo así el mapa cromosómico.

En el mapa de ligamiento de un cromosoma, las distancias genéticas guardan una relación con las distancias físicas (número de pares de bases), pero esa relación no es en absoluto directa. Así, por ejemplo, hemos encontrado que la frecuencia de recombinación entre un determinado par de marcadores suele diferir significativamente entre individuos de distinto sexo. Esto es, para dos marcadores concretos, la probabilidad de que se recombinen durante la meiosis es muy distinta según el cromosoma que los porta se haya heredado de la madre o del padre. En el cromosoma 13, por ejemplo, las frecuencias de recombinación son varias veces mayores en las hembras. En el cromosoma 11 ocurre lo contrario en cierto segmento; en otro segmento adyacente a ése los dos sexos presentan frecuencias de recombinación similares. Se ignoran las bases moleculares de esas extrañas variaciones. A efectos operativos hemos preparado dos mapas, uno para cada sexo, con idénticas secuencias de marcadores, pero con distancias genéticas distintas.

Tenemos ya una primera versión de los mapas de la mayoría de los cromosomas humanos. Otro grupo de investigadores ha publicado recientemente una colección similar de mapas, basados en un número menor de familias. Con frecuencia ocurre que los marcadores estudiados y desarrollados en un laboratorio complementan a los cartografiados en otro. Ello permite, a veces, rellenar los "huecos" existentes entre grupos de marcadores ligados.

Jean Dausset ha organizado, en el Centro de Estudios del Polimorfismo Humano (CEPH), con sede en París, la infraestructura necesaria para facilitar la cooperación entre los grupos que trabajan en el tema. El CEPH recoge, mantiene y distribuye ADN de unas 40 familias. La colección se nutre fundamentalmente de nuestras familias de Utah, aunque también incluye ADN de familias estudiadas por otros grupos. Investigadores de todo el mundo (nosotros también) pueden obtener juegos completos de ADN de dicha colección. Ello compromete a informar al CEPH de los marcadores y pautas de herencia estudiados. El centro, a su vez, ofrece dichos datos a todos los investigadores, con lo que, poco a poco, se van sentando las bases para la construcción de un único mapa genético.

En los próximos años, con la consecución de un mapa de alta resolución, el genoma humano dejará definitivamente de ser un territorio inexplorado. Se espera que dicho mapa permita la localización de los genes responsables de la mayoría de las enfermedades genéticas bien caracterizadas que quedan por cartografiar. Un mapa de ligamiento completo servirá también para abordar otra investigación a gran escala del genoma, aún en una etapa de planificación: la determinación de la secuencia completa de bases del ADN humano. Es probable que primero se secuencien pequeños "islotes" de ADN, repartidos a lo largo de todos los cromosomas. Los grupos de ligamiento presentes en cada uno de esos islotes servirán para situarlos adecuadamente en el genoma.

Un defecto en un espejo universal

Si no fuera por una ligera asimetría en un "espejo" llamado invariancia CP, no existiría el universo que conocemos; antes bien, se hallaría desprovisto de materia. ¿Qué fuerza de la naturaleza provoca la violación de la simetría CP?

Robert K. Adair

¿Por qué hay materia en el universo? Si, en vez de aproximada, la simetría que se ha observado entre materia y antimateria fuera perfecta, el universo presentaría una elegante sencillez, aunque carecería de materia y de las criaturas formadas por esta materia que pudiesen admirar su elegancia. Nuestro universo proviene de un defecto en una simetría que ofrece un "espejo" universal llamado espejo *CP*: una simetría que establece la invariancia en el resultado de algunos sucesos de la naturaleza si se cambia la materia por antimateria (*C*) y se mira el suceso en un espejo (*P*).

Parece probable que, en algún momento, antes de que se cumpliera la primera millonésima de segundo desde que el universo naciera en la bola de fuego de la gran explosión ("big bang"), hubiera la misma cantidad de materia que de antimateria. Habría casi exactamente el mismo número de partículas que de antipartículas, todas ellas en equilibrio termodinámico bajo unas condiciones de presión y temperatura enormes. (Para cada partícula existe una antipartícula, dotada de masa idéntica y propiedades eléctricas opuestas; el antielectrón o positrón, por ejemplo, es la antipartícula del electrón.) A medida que el universo se fue expandiendo y enfriando, la mayoría de partículas encontraron sus correspondientes antipartículas y se fueron aniquilando entre sí. Si la simetría *CP* fuera exacta, sólo habrían quedado las escasísimas partículas y antipartículas que por casualidad no hubieran encontrado pareja para aniquilarse. Pero se produjo un leve fallo en la simetría y vino así a salvarse un protón de cada mil millones y un electrón de cada mil millones. Andando el tiempo, se formarían, por esa razón, las galaxias, las estrellas, los planetas y los seres humanos.

A pesar de que esta violación de la invariancia *CP* constituye el núcleo del enigma de nuestra existencia, y a pesar del estudio teórico empeñado y los ensayos experimentales dedicados a *CP*, no sabemos de la asimetría *CP* mucho más que lo que se sabía veinticinco años atrás, cuando se descubrió dicha anomalía. Sin embargo, en los últimos tiempos han ido apareciendo nuevas ideas cuya profundización ha urgido una nueva generación de experimentos —uno de los cuales emprendimos mis colaboradores y yo en el Laboratorio Nacional de Brookhaven— que prometen arrojar nueva luz sobre el defecto. Los experimentos permitirán examinar, así se espera, el carácter de la fuerza responsable de la violación de la simetría *CP* y determinar si la minúscula asimetría se debe a una fuerza débil (milidébil) que actúa dos veces o a una fuerza muy débil (superdébil) que actúa una sola vez en las interacciones entre partículas.

Una plena apreciación de la leve, aunque crucial, asimetría del espejo *CP* exige acercarse a la función que cumple la simetría en la construcción del universo. La mayoría de los físicos participa de la opinión de que debe haber una gran ecuación que defina las propiedades de las fuerzas y las partículas elementales, ecuación que expresaría las simetrías fundamentales del universo. Aunque se desconocen todavía los pormenores de la estructura de la ecuación, se dominan ya muchas de sus simetrías más importantes, cuya intelección no encierra mayor dificultad.

Tal como demostró a principios de siglo Amalie Emma Noether, de la Universidad de Göttingen, las simetrías de las ecuaciones fundamentales suelen estar relacionadas con leyes de conservación. Así, las observaciones de que el momento, la energía y el momento angular se conservan, demuestran que la gran ecuación no distingue entre lugares, tiempos y direcciones. Resulta inmediato interrogarse acerca de otras simetrías. En concreto, ¿distingue la ecuación el giro levógiro del dextrógiro (imágenes especulares) y define por tanto el sentido de avance de un tornillo? ¿Distingue la ecuación entre partículas y antipartículas?

Los experimentos de laboratorio pueden ofrecer la respuesta a estas cuestiones. Ilustraré el carácter de la investigación de las simetrías fundamentales con la ayuda de la intrépida heroína de Lewis Carroll, Alicia. Y empezaremos con un estudio de la simetría bajo las reflexiones en un espejo de tocador; de una manera más formal lo llamaríamos espejo *P*. (La abreviatura *P* remite a paridad.) ¿Puede Alicia, despertándose quizá de un sueño de Carroll, decir si se encuentra en el mundo que hay detrás del espejo o en su propio mundo? Alicia sólo saldrá de dudas si da con algún proceso o alguna estructura fundamental que definan la dirección de avance de un tornillo. Un tornillo orientado a derechas en su mundo será un tornillo orientado a izquierdas en el espejo.

Hasta 1956 los físicos hubieran respondido que Alicia se perdería entre el mundo real y el mundo especular. Por la sencilla razón de que sosteníase entonces que ninguna de las interacciones fundamentales —la gravedad, el electromagnetismo, las interacciones fuertes (que son responsables de las fuerzas nucleares) y las interacciones débiles (que son responsables de ciertas clases de desintegraciones nucleares)— definen la dirección de avance del tornillo. Con otras palabras, creíase que el espejo *P* era simétrico y no había forma de distinguir el resultado de una reacción entre partículas elementales del

resultado de la imagen especular de dicha reacción.

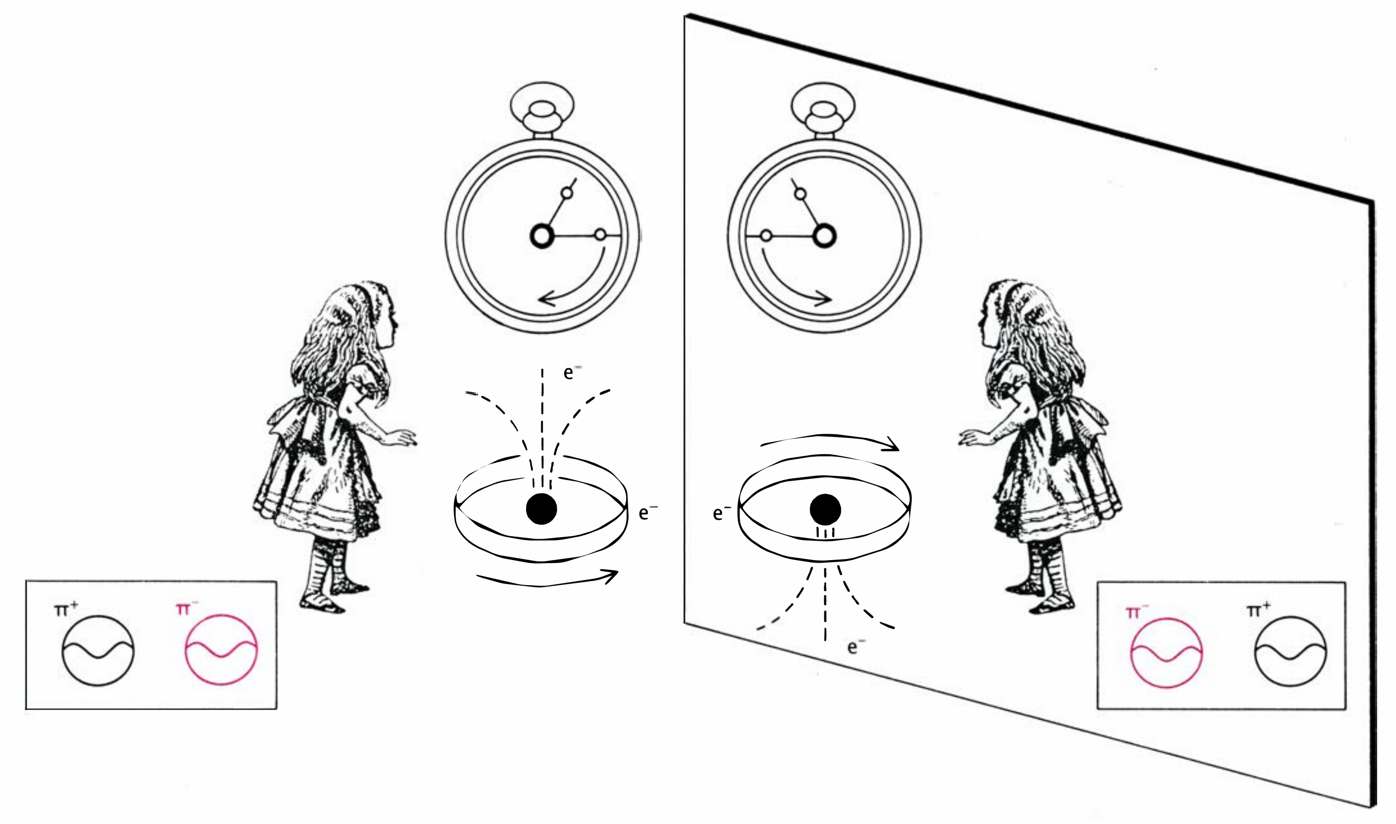
Así las cosas, Richard H. Dalitz, que trabajaba en la Universidad de Cornell, demostró en aquellos días que las propiedades de desintegración de una partícula llamada mesón *K* a través de la interacción débil parecía contradecir la idea según la cual la naturaleza no distingue entre la orientación dextrógira y la orientación levógira. En el curso del proceso de estudio de este asunto, Tsung Dao Lee y Chen-Ning Yang probaron, en 1956, que aceptar la hipótesis de que las interacciones débiles definen una dirección de avance del tornillo no implica ninguna contradicción. En esencia, la teoría de Lee y Yang, que les hizo acreedores del premio Nobel en 1957, muestra que el espejo *P* puede ser asimétrico con respecto a las interacciones débiles. En diciembre de 1956, Chien Shiung Wu, de la Universidad de Columbia, Ernest Ambler, de la Oficina Nacional de Pesos y Medidas, y otros colaboradores llevaron a cabo un experimento sugerido por Lee y Yang. El ensayo corroboró que las interacciones débiles definían una dirección de avance del tornillo: la naturaleza distinguía la orientación dextrógira de la orientación levógira.

En resumen, Alicia puede averiguar dónde se halla, si en el mundo del espejo o en su mundo real, con un experimento de las características del realizado por Wu y Ambler. Basta con que coloque un cilindro de cobalto 60, un isótopo del cobalto que se desintegra emitiendo electrones a través de las interacciones débiles, en un campo magnético generado por electrones que circulen en sentido contrario a las agujas del reloj por una espira circular [*véase la parte izquierda de la figura 1*]. Según leyes elementales de la física, el campo magnético generado por los electrones que se mueven en la espira está dirigido hacia abajo. Pero los núcleos de todos los átomos tienen un cierto espín, o momento angular, que les convierte en pequeños imanes que giran en torno a sus ejes. En el caso de los núcleos de cobalto, el campo magnético alinea los espines nucleares hacia abajo, según la dirección del campo.

En su laboratorio del mundo real, Alicia observa que los electrones emitidos por el cobalto a través de la desintegración débil se dirigen hacia arriba, en el sentido opuesto a los espines nucleares. Al otro lado del espejo, Alicia [*véase la parte derecha de la figura 1*] lleva a cabo el mismo experimento; ahora bien, como los electrones de la espira en el mundo especular se mueven en el sentido de las agujas del reloj, generan un campo magnético dirigido hacia arriba que alinea en ese sentido los espines de los núcleos de cobalto. Por tanto, los electrones emitidos por los núcleos de cobalto en el mundo del espejo se dirigen hacia abajo. Lo que significa que el mundo del espejo no es una reflexión absoluta del mundo real. Se rompe la simetría del espejo *P*.

Si Charles Dodgson hubiera nacido un siglo más tarde, podría haberle proporcionado otro espejo a Alicia con el que disfrutar nuevas aventuras. Quizás ella podría haber pasado entonces a través de un espejo *C*, que transformaría partículas en antipartículas. (La *C* se refiere a la conjugación de carga.)

A finales de la década de 1920 el físico británico P. A. M. Dirac, que postuló la existencia de la antimateria, de-



1. ALICIA EN EL PAIS DE LAS MARAVILLAS: la representación de su problema nos facilita la comprensión de las leyes de la simetría en la naturaleza. Se ilustra aquí una prueba de la simetría de reflexión en un espejo ordinario, o espejo *P* según se le denomina de una manera formal. (La *P* abrevia el término paridad.) ¿Puede distinguir Alicia si se encuentra en el mundo real (*izquierda*) o en el mundo especular (*derecha*)? Para averiguarlo, coloca un cilindro de cobalto 60, un isótopo del cobalto que se desintegra emitiendo electrones, en el campo magnético generado por unos electrones que circulan en sentido contrario a las agujas del reloj alrededor de una espira circular. El campo apunta hacia abajo y alinea los espines (momentos angulares) de los núcleos de cobalto hacia abajo. Los electrones emitidos salen hacia arriba. En el mundo del espejo, los electrones circulan por la espira en el sentido de las agujas del reloj, de manera que el campo magnético apunta hacia arriba. En el mundo del espejo, los espines de los núcleos se alinean hacia arriba y los electrones emitidos se dirigen hacia abajo. Como los electrones emitidos viajan en sentidos opuestos en uno y otro mundo, la reflexión no es perfecta: la simetría del espejo *P* está rota. El espejo *P* conserva las identidades de las partículas y las antipartículas (*recuadros rectangulares*). Una antipartícula (aquí el mesón π^- cargado negativamente) tiene la misma masa que su correspondiente partícula (en este caso será el mesón π^+, cargado positivamente) pero sus propiedades eléctricas son opuestas.

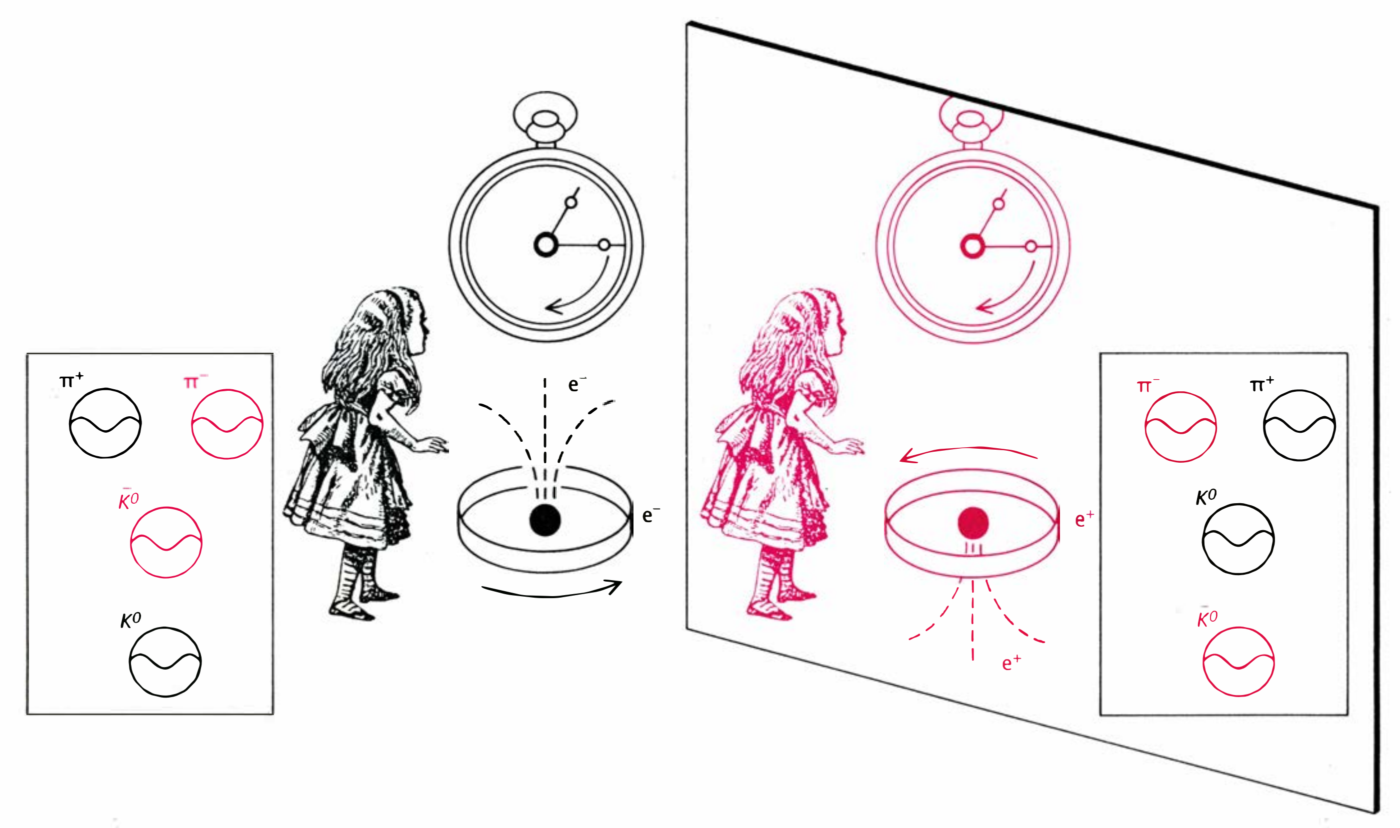
mostró que la materia y la antimateria debían comportarse de la misma manera bajo la acción de las fuerzas electromagnéticas. Más tarde se vio que también la gravedad y las fuerzas nucleares fuertes producen idénticos resultados al actuar sobre materia y antimateria. Si la simetría vale para todas las interacciones, Alicia no podrá realizar ninguna observación que le indique si ha pasado o no a través del espejo *C*.

Sin embargo, el trabajo experimental y teórico relacionado con el descubrimiento de la asimetría del espejo *P* también mostró que las interacciones débiles distinguían entre materia y antimateria: la simetría del espejo *C* quedaba rota también. En concreto, los núcleos de anticobalto alineados emiten positrones (denominación más corriente que antielectrones) preferentemente en el sentido de alineamiento.

Así, pues, Alicia puede saber si ha pasado o no a través del espejo *C* determinando el sentido de emisión de los electrones liberados por los núcleos de cobalto alineados o, de forma alternativa, determinando el sentido de emisión de los positrones liberados por los núcleos de anticobalto alineados. De nuevo, en su mundo real los electrones que viajan en sentido contrario a las agujas del reloj por una espira generan un campo magnético dirigido hacia abajo; los núcleos de cobalto se alinean hacia abajo y los electrones que emiten van hacia arriba [*véase la mitad izquierda de la figura 2*]. En el antimundo visto a través del espejo *C*, los positrones que viajan por la espira se mueven también en sentido contrario a las agujas del reloj (ya que el espejo *C* invierte la carga); ahora bien, como su carga es del signo opuesto a la de los electrones, el campo magnético está dirigido hacia arriba. Los núcleos de anticobalto tienen propiedades magnéticas opuestas a las de los núcleos de cobalto ordinario, de manera que sus espines apuntarán hacia abajo. Por último, los positrones emitidos por los núcleos de anticobalto salen en el sentido que marcan los espines nucleares, en este caso hacia abajo [*véase la mitad derecha de la figura 2*]. El hecho de que los positrones emitidos salgan en sentido opuesto al seguido por los electrones emitidos indica a Alicia que ha pasado al antimundo a través del espejo *C*.

Hemos aludido antes a la importancia que los físicos conceden al hallazgo de simetrías en la naturaleza. Cabe, pues, imaginar con qué sorpresa muchos físicos recibieron la asimetría del espejo *P*. A mí me produjo la sensación de que jamás podría considerar cierto nada de lo que sabía. La tierra firme que pisaba se había convertido en arena movediza. Las aguas volvieron a su cauce y el universo tornaba a adquirir buena parte del orden que creíamos perdido con la observación de que la asimetría del espejo *P* parecía siempre estar acompañada por una asimetría compensatoria en el espejo *C*, de manera que todo quedara simétrico bajo *CP*. En otras palabras, si Alicia pasaba a través del espejo *CP*, hallaría los mismos resultados que si estuviera en su mundo real [*véase la figura 3*]. Durante unos siete años se mantuvo la idea según la cual las fuerzas elementales eran absolutamente simétricas bajo *CP*.

Llegamos a 1964. James Cronin y Val L. Fitch, de la Universidad de Princeton, que trabajaban en Brookhaven, demostraron que se producían pequeñas violaciones de la simetría *CP*. Me ocuparé de su experimento más adelante; estudiaremos ahora un resultado afín, de explicación física más sencilla. Si Alicia acelera a altas energías un haz de mesones K^+ cargados positivamente y dirige las partículas hacia un bloque



2. EL ESPEJO *C* cambia partículas por antipartículas y antipartículas por partículas; mantiene, sin embargo, las orientaciones relativas de los objetos que "refleja". (La letra *C* designa conjugación de carga.) A la izquierda, los electrones que circulan en sentido antihorario por una espira producen un campo magnético dirigido hacia abajo; los núcleos de cobalto del cilindro emiten electrones que se dirigen hacia arriba. En el antimundo visto a través del espejo *C* (*derecha*), los positrones, o antielectrones, viajan también en sentido antihorario a través de la espira (el espejo *C* sólo cambia la carga); ahora bien, por tener los positrones una carga opuesta a la de los electrones, el campo magnético apunta hacia arriba. Los núcleos de anticobalto tienen propiedades magnéticas opuestas a las de los núcleos de cobalto ordinario, de manera que sus espines se alinean hacia abajo. Los núcleos de anticobalto emiten positrones en la dirección de sus espines nucleares: y así, los positrones se dirigen hacia abajo. Puesto que los positrones emitidos viajan en dirección opuesta a la de los electrones emitidos, la simetría del espejo *C* se rompe. Los recuadros muestran que el espejo *C* cambia la carga de los mesones π^+ y π^- y transforma el mesón K^0 en $\bar{K}^0$.

de cobre, aparecerá un haz de mesones K neutros. Lejos del blanco, los mesones K neutros se desintegran en positrones de forma algo más frecuente que en electrones. A través del espejo CP, Alicia acelera un haz de mesones K^- cargados negativamente (las antipartículas de los mesones K^+) contra un bloque de anticobre. De nuevo emergen mesones K neutros que también se desintegran en positrones con una frecuencia un poco mayor que la de los electrones. En ambos casos, los mesones K neutros se desintegran en positrones con preferencia a los electrones; por tanto, la simetría CP se rompe y Alicia puede determinar si está en su mundo real o ha pasado a través de los espejos P y C.

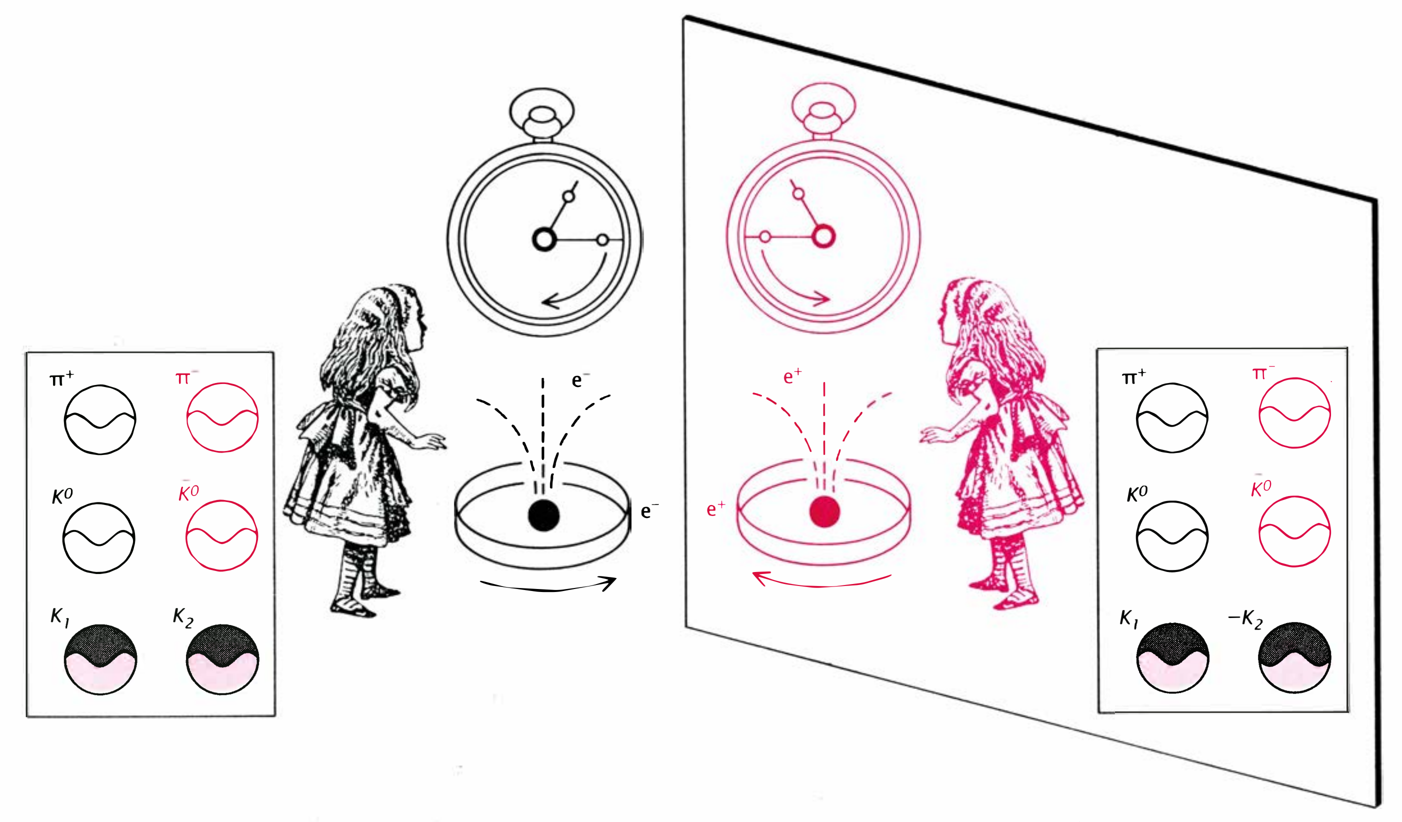
Para llevar a cabo tal determinación, Alicia se basa en que un campo magnético desviará toda partícula cargada que se mueva a través del mismo. En su análisis, lo que hará será medir la desviación de los positrones —de carga positiva— y de los electrones —de carga negativa— en que se desintegran los mesones K neutros. De nuevo, en su mundo real el campo magnético está producido por electrones que circulan en sentido contrario a las agujas del reloj a través de una espira circular; el campo apunta, pues, hacia abajo. Al otro lado del espejo CP, el campo magnético lo producen los positrones que circulan en sentido horario; allí, el campo apunta también hacia abajo. Alicia se encuentra con que los productos de desintegración más abundantes provenientes de los mesones K neutros (los positrones) se desvían hacia ella, mientras que en el mundo invertido por CP se encuentra con que los productos de desintegración se desvían alejándose de ella [*véase la figura 4*]. Observando la desviación puede saber dónde está. El experimento muestra que no todo es invariante tras la combinación de la reflexión en un espejo P y el cambio de partícula por antipartícula producido por un espejo C.

El experimento demuestra una diferencia, profunda y fundamental, entre la materia y la antimateria; asimismo, elimina la prohibición de que el universo presente un exceso de materia, norma que viene impuesta por la simetría CP. Aún no conocemos el carácter exacto de los mecanismos que conducen al exceso de materia de que consta el mundo, pero sabemos, a partir del experimento de Alicia, que nuestra existencia no está en contradicción con las simetrías que observamos.

Profundizar en el entendimiento de la asimetría CP exige conocer con más detalle las partículas sujetas a la asimetría. ¿Por qué los mesones K neutros se comportan así?

Hasta ahora, por criterios de sencillez, me he referido a los mesones K neutros como si se tratara de una clase sola de partículas. En realidad las cosas son más complicadas. En primer lugar, existe el mesón K neutro denominado K^o. Está luego la antipartícula de dicha partícula, el mesón $\bar{K}^o$. La barra sobre la K indica antipartícula.

Un mesón K neutro no tiene por qué ser un mesón K^o o $\bar{K}^o$. Puede ser una mezcla de los dos. Algo que parece un concepto extraño si uno imagina los mesones en términos corpusculares, resulta menos curioso si nos atenemos a la descripción que en el contexto de la mecánica cuántica propone De Broglie: nos lo podemos representar como ondas, con propiedades muy parecidas a las ondas de agua o las luminosas. De la misma manera que se pueden producir nuevas ondas mezclando amplitudes de dos ondas distintas, se pueden construir también nuevos mesones K neutros a partir de las amplitudes de los mesones K^o y $\bar{K}^o$.



3. COMBINACION DE UN ESPEJO C Y UN ESPEJO P para formar un espejo CP. Los físicos creyeron, hasta el año 1964, que dicha combinación estaba libre de defectos. Se pensaba que la asimetría del espejo P iría siempre acompañada de la asimetría del espejo C, de manera que un espejo CP sería simétrico. En otras palabras, si Alicia tuviera que atravesar el espejo CP, los resultados de cualquier experimento que realizara serían los mismos. En el caso que se ilustra aquí, los resultados son efectivamente los mismos: tanto los electrones (*izquierda*) como los positrones (*derecha*) son emitidos hacia arriba. Los recuadros de abajo indican el efecto del espejo CP sobre varias partículas, incluyendo los mesones de carga neutra K_1 y K_2, cada uno de los cuales está constituido por iguales cantidades de los mesones K^o y $\bar{K}^o$. (La "mezcla" de partículas es uno de los extraños fenómenos predichos por la mecánica cuántica.) Cada partícula tiene asociada a ella una función de ondas, con propiedades muy parecidas a las de una onda de agua o una onda luminosa, que describe las características de la partícula. La función de ondas asociada al mesón K_1 es simétrica con respecto al espejo CP y la función de ondas del mesón K_2 es antisimétrica (de ahí el signo menos).

Se pueden producir mesones K neutros mezclando mesones K^0 y $\bar{K}^0$ en cualquier proporción; ahora bien, para nuestro propósito, necesitamos sólo considerar el mesón K_1, que se forma sumando amplitudes iguales de K^0 y $\bar{K}^0$ y el mesón K_2, que se forma restando una amplitud K^0 de una amplitud $\bar{K}^0$ igual. Los mesones K_1 y K_2 vendrían a ser mitad partícula y mitad antipartícula.

No hay ninguna razón a priori para pensar que los mesones K^0 y $\bar{K}^0$ sean partículas "puras" y mezclas los mesones K_1 y K_2. Igualmente, podríamos tomar por base los mesones K_1 y K_2 y obtener los mesones K^0 y $\bar{K}^0$ como mezclas. El mesón K^0 es la suma del K_1 y el K_2 ($K^0 = K_1 + K_2$), y el mesón $\bar{K}^0$ es la diferencia ($\bar{K}^0 = K_1 - K_2$).

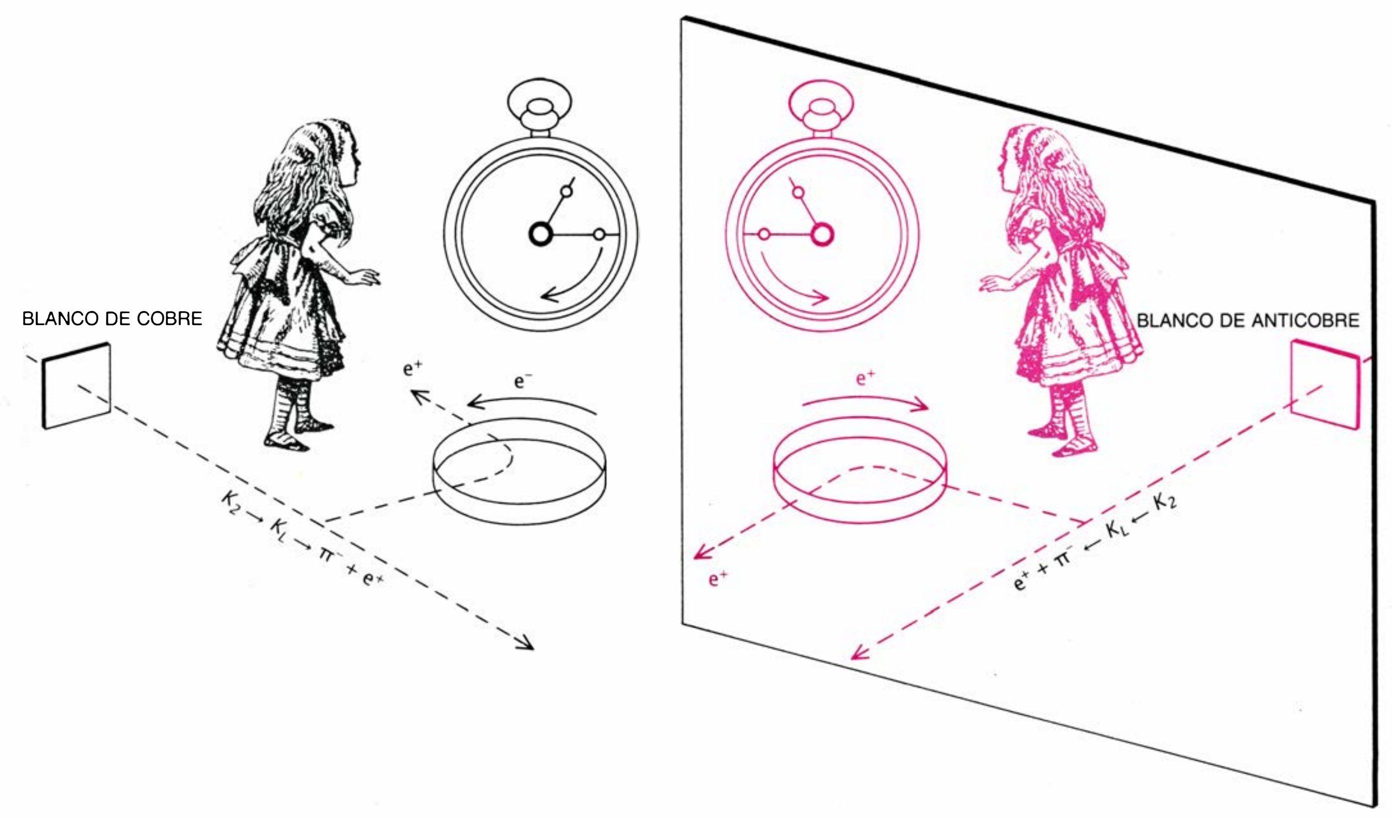
Cuando Alicia bombardeaba su blanco de cobre con mesones K^+ cargados positivamente, emergían mesones K^0 de carga neutra. En el mundo CP, al bombardear su blanco de anticobre con mesones K^- cargados negativamente, resultaban mesones $\bar{K}^0$ de carga neutra. Ya he indicado que cada uno de estos haces iniciales de mesones K^0 y $\bar{K}^0$ se puede representar convertidos en una mezcla con cantidades iguales de mesones K_1 y K_2. La componente K_1 de ambos haces se desintegra bastante rápidamente en pares de partículas llamadas mesones π, quedando sólo la componente K_2.

Como la parte K^0 del mesón K_2 se desintegra en positrones y la parte $\bar{K}^0$ (presente en igual cantidad) se desintegra en electrones, de la desintegración del K_2 no resulta un exceso de positrones sobre electrones; una observación de dicha desintegración no puede, pues, indicar a Alicia dónde se encuentra. Sin embargo, el efecto del espejo CP que da lugar a la violación de la invariancia actúa agregando espontáneamente una muy pequeña parte de K_1 en el K_2 para dar lugar a una combinación llamada K_L; con una parte ligeramente mayor de K^0 que de $\bar{K}^0$, K_L se desintegra más a menudo en positrones que en electrones. (La L remite a vida larga.) Este efecto de la violación de la invariancia CP permitió a Alicia saber dónde se hallaba.

La adición de una parte de K_1 al K_2 —el hecho que viola la invariancia CP— se pone también de manifiesto de otra manera. Tal como observaron Cronin y Fitch en 1964, el K_L se desintegra en dos mesones π. Esta desintegración proviene de la parte K_1 que hay en el K_L; K_2 no se desintegra casi nunca en dos mesones π.

La distinta desintegración de los mesones K_1 y K_2 radica en las propiedades CP de las partículas. Bajo la reflexión en un espejo CP perfecto, la amplitud de onda del K_1 no cambia, y lo mismo le sucede a la amplitud de la onda que describe dos mesones π, pero la amplitud del K_2 queda invertida y cambia de signo. Con ello, el K_1 se desintegra rápidamente en pares de mesones π^+ y π^- y en pares de mesones π^0. Pero no ocurre lo mismo con K_2, que se desintegra de otras maneras, aunque despacio.

Igual que Alicia, Cronin y Fitch empezaron con un haz de mesones K^0 [*véase la figura 6*]. De nuevo, la componente K_1 del haz se desintegró rápi-



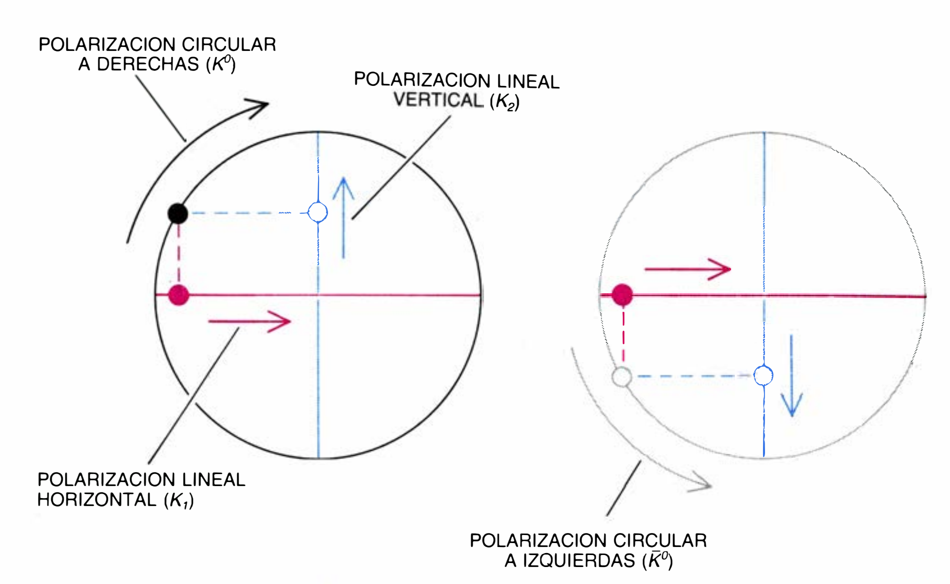
4. ALICIA DESCUBRE EL DEFECTO DEL ESPEJO *CP*. Ella acelera un haz de mesones de carga positiva K^+ dirigiéndolos hacia un bloque de cobre (*izquierda*); se produce entonces un haz de mesones de carga neutra K^0. La componente K_1 de los mesones K^0 se desintegra en pares de mesones π, quedando sólo la componente K_2, que después de viajar cierto tiempo se convierte en un mesón K_L mediante la adición de una minúscula cantidad de componente K_1. Esta combinación K_L se desintegra en positrones más frecuentemente que en electrones. Mientras tanto, Alicia, al otro lado del espejo *CP*, acelera un haz de mesones de carga negativa K^- hacia un bloque de anticobre (*derecha*), produciendo un haz de mesones de carga neutra $\bar{K}^0$. La componente K_1 de los mesones $\bar{K}^0$ se desintegra en pares de mesones π, quedando de nuevo la componente K_2. El mesón K_2, después de recorrer cierta distancia, se convierte en el mesón K_L, que, a su vez, se desintegra en positrones con más frecuencia que en electrones. En ambos casos, las partículas que eran mesones K_2 neutros se desintegran en positrones con más frecuencia que en electrones; por tanto, la simetría *CP* se rompe. Alicia puede determinar si ha pasado a través del espejo *CP* sabiendo que un campo magnético desviará una partícula cargada que se mueva perpendicularmente a través del campo. En el mundo real, produce otra vez un campo magnético haciendo circular electrones en sentido antihorario a través de una espira, de manera que el campo apunta hacia abajo. Al otro lado del espejo *CP*, los positrones que circulan en sentido horario generan un campo que también apunta hacia abajo. Alicia descubre que los positrones de los mesones K_2 se desvían hacia ella, mientras que la Alicia del otro lado del espejo *CP* se da cuenta de que los positrones se alejan de ella. La simetría del espejo *CP* también se rompe.

damente en pares de mesones π, quedando sólo la componente K_2. Si el espejo CP fuera perfecto, el mesón K continuaría siendo un mesón K_2 y no se desintegraría en pares de mesones π. Sin embargo, la asimetría CP actúa de manera que K_2, espontáneamente, adquiere una minúscula cantidad de amplitud K_1, debido a la cual aproximadamente uno de cada quinientos mesones K_L se desintegra en un par de mesones π.

¿Se rompe el espejo CP con sistemas en los que no intervenga el mesón K? Aunque los físicos están bastante convencidos de que el defecto del espejo CP es universal, da la impresión de que la magnitud de los efectos resultó demasiado pequeña para llegar a observarlos en otros sistemas que se han investigado. Quizá sea, por tanto, una nueva fuerza muy débil la responsable del efecto. ¿Qué carácter tendría esta fuerza?

He indicado que la violación CP observada en mesones K puede describirse en términos de la adición de una pequeña parte de K_1 al mesón K_2. Dicho de manera algo distinta, una pequeña parte del mesón $\bar{K}^o$, que forma parte del K_2, pasa a ser un mesón K^o.

Según es conocido, los mesones constan de un quark y de un antiquark. Los mesones K están formados por dos clases de quarks, los denominados abajo ("down") y extraño ("strange") con sus correspondientes antiquarks: antiabajo y antiextraño; el mesón K^o consta de un quark abajo y un quark antiextraño; el mesón $\bar{K}^o$, por su parte, consta de un quark extraño y un quark antiabajo. Para que el mesón $\bar{K}^o$ se convierta en mesón K^o deben ocurrir dos cosas: que un quark antiabajo se convierta en un quark antiextraño y que un quark extraño se mude en un quark abajo. (Los quarks no se transforman en antiquarks.) Se sabe que son las fuerzas débiles las que inducen tales cambios de quarks, pero conservan CP, de manera que un cambio en un sentido (de $\bar{K}^o$ en K^o) no es más probable que el inverso (de K^o en $\bar{K}^o$). No hay cambio neto: el espejo CP no se rompe.



5. ANALOGIA entre las propiedades de la luz y los mesones K neutros. Puede ayudarnos a explicar la "mezcla" de dichos mesones. La luz puede polarizarse de múltiples formas. Pensemos en un haz de luz que salga de esta página. Si los campos eléctricos oscilantes de dicho haz apuntan a derecha e izquierda del plano de la página, se dice que la luz está polarizada horizontalmente; se puede considerar que al mesón K_1 le corresponde tal polarización. Si los campos eléctricos se dirigen hacia arriba y abajo del plano de la página, el haz está polarizado verticalmente; al mesón K_2 le corresponde esa polarización. Si la dirección del campo eléctrico se mueve circularmente, como si de la dirección de la hendidura de un tornillo se tratara, que cambia a medida que éste avanza al enroscarse, se dice que la luz está polarizada circularmente. Si la dirección del campo varía en el sentido definido por el avance de un tornillo orientado a derechas, la luz está polarizada circularmente a derechas, lo que corresponde al mesón K^o; si la dirección varía como un tornillo orientado a izquierdas, la luz está polarizada circularmente a izquierdas, y esto representa al mesón $\bar{K}^o$. La polarización circular a derechas es la suma de la polarización lineal horizontal y la polarización lineal vertical (*izquierda*), mientras que la polarización circular a izquierdas es la diferencia (*derecha*). Prosiguiendo con la misma analogía, la amplitud de onda del mesón K^o es la suma de las amplitudes K_1 y K_2 (de manera que en cierto sentido $K^o = K_1 + K_2$), mientras que la amplitud de onda del mesón $\bar{K}^o$ es la diferencia ($\bar{K}^o = K_1 - K_2$).

Sin embargo, la existencia de una fuerza asimétrica respecto a CP podría inducir la violación observada de la invariancia CP. Esta fuerza debería cambiar un quark antiabajo en un quark antiextraño con más frecuencia de la que cambia un quark antiextraño en un quark antiabajo; debería, asimismo, cambiar un quark extraño en un quark abajo con frecuencia mayor que un quark abajo en un quark extraño. Un mesón $\bar{K}^o$ muy raramente se convierte en un mesón K^o, lo que da pie a pensar que la transición puede ser una coincidencia casual de dos procesos independientes más o menos raros. La violación de CP podría ser producida por la coincidencia de una transición simétrica respecto a CP, moderadamente improbable, generada por la fuerza débil, y una transición asimétrica respecto a CP, algo más improbable (en un factor alrededor de 500) inducida por una fuerza asimétrica bajo CP. Una fuerza moderadamente débil de este tenor se denomina milidébil (por ser unas 1000 veces más débil que la fuerza débil).

La familia de mesones K neutros constituye, pues, un sistema que proporciona pruebas bastante nítidas de la violación de la invariancia CP. ¿Existe algún otro sistema así? Trabajos recientes, sobre todo los realizados en el acelerador electrón-positrón DESY de Hamburgo, han despertado un notable interés por el sistema de mesones B neutros, que guarda importantes paralelismos con el sistema de mesones K. De la misma manera que los mesones K neutros están constituidos por quarks abajo y antiextraños o quarks antiabajo y extraños, los mesones B están formados por quarks abajo y antifondo o quarks antiabajo y fondo. Los quarks fondo son similares a los quarks extraños aunque mucho más pesados. Por tanto, los mesones B son muy parecidos a los mesones K con la salvedad de que pesan unas 10 veces más. Creo que la mayor masa del mesón B, entre otros factores, actúa de tal manera que no me parece previsible que en un futuro próximo vayamos a sacar mucho en claro acerca de la violación de la invariancia CP a partir del estudio de los mesones B.

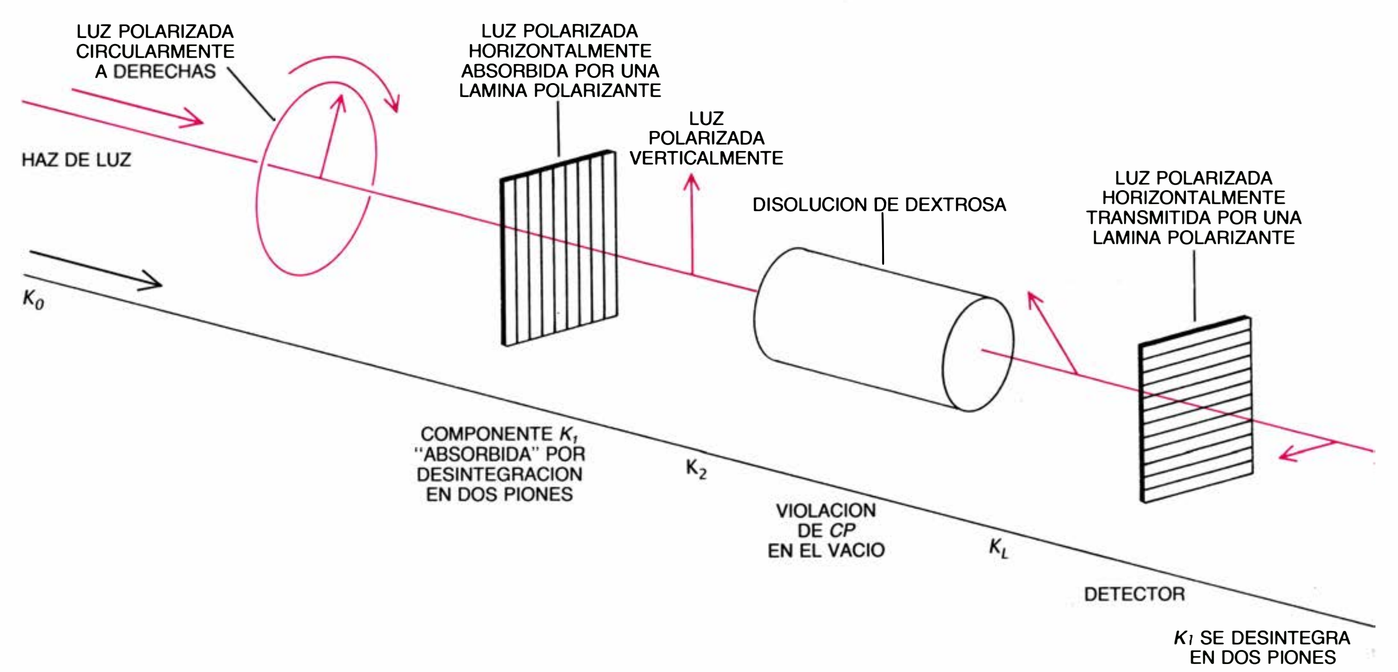
La hipótesis de la fuerza milidébil, aunque atractiva, presenta un inconveniente: quizá no sea lo bastante débil. ¿Por qué la observación de la violación CP ha estado limitada al mesón K neutro? De una fuerza asimétrica respecto a CP tan fuerte como la milidébil podría esperarse alguna otra manifestación. ¿Podría dar cuenta del defecto del espejo CP una fuerza mucho más débil? Tal vez. Lincoln Wolfenstein ha sugerido que, en vez de tratarse de la coincidencia de una transición débil y una transición milidébil, la violación de la invariancia CP podría estar generada por una fuerza debilísima que produjera ambos cambios actuando una sola vez. Resultaría muy difícil observar, en ninguna otra parte, esta fuerza, superdébil, según se la califica, mil millones de veces más débil que la fuerza débil.

Ambas nociones, de fuerza milidébil y fuerza superdébil, son agrupaciones de teorías más que teorías sueltas. La mayor parte de la investigación sobre *CP* se ha encaminado hacia la detección de efectos milidébiles. La búsqueda no ha dado frutos positivos, pero hasta hace poco los resultados negativos de los experimentos no podían aprovecharse para excluir modelos milidébiles, ya que no existía una predicción general de las consecuencias de los modelos. Se ha apuntado ya una característica bastante específica de la fuerza milidébil en trabajos teóricos iniciados por Frederick J. Gilman, de la Universidad de Stanford, y Mark Wise, del Instituto de Tecnología de California, luego desarrollados por otros. Según esta teoría, a consecuencia de las fuerzas milidébiles asimétricas bajo *CP* el mesón K_2 no sólo debería transformarse parcialmente en el mesón K_1, sino también desintegrarse directamente en dos mesones π.

La desintegración directa del mesón K_2 en pares de mesones π diferiría de la desintegración del mesón K_1 en pares de mesones π y debiera, por tanto, detectarse. La probabilidad de que un mesón K_1 se desintegre directamente en pares de mesones π^+ y π^- es el doble que la de desintegrarse en un par de mesones π^0. Esta razón debe invertirse en el caso del mesón K_2: debería ser dos veces más probable que se desintegrara en pares de mesones π^0. En ausencia de desintegraciones directas de K_2, el mesón K_L se desintegrará en dos piones sólo a través de su parte K_1 y presentará, por tanto, la razón de carga de los dos piones propia del K_1. Si también hubiera desintegraciones directas en dos piones de la parte K_2 del K_L, la razón quedaría modificada. Aunque las diferentes teorías sobre la asimetría *CP* predicen variaciones, casi todas conducen a modificaciones sustanciales de la razón dos a uno, propia de las desintegraciones K_1 puras; la magnitud general del efecto debe ser del orden del 30 por ciento

Los primeros experimentos parecían excluir diferencias de ese rango, del 30 por ciento. Sin embargo, algunos modelos teóricos predicen efectos mucho menores. Mediciones más recientes realizadas en Brookhaven por un grupo de la Universidad de Yale y de Brookhaven que dirigieron Michael P. Schmidt, William M. Morse y el autor del artículo, y otras en el Laboratorio del Acelerador Nacional Fermi por un grupo de la Universidad de Chicago, la Universidad de Wisconsin y el Centro de Investigación Nuclear de Saclay, dirigido por Cronin y Bruce D. Winstein, sugieren que la diferencia es inferior al 10 por ciento. Estos resultados han forzado la exclusión de algunos interesantes modelos milidébiles sobre la violación de la invariancia respecto a *CP*, aunque no todos. Un experimento modificado del Fermilab y otras mediciones que se están realizando en el CERN, el laboratorio europeo de física de partículas, abrigan la esperanza de alcanzar mayores precisiones. De hecho, el grupo del CERN ha anunciado resultados provisionales que sugieren una variación del 2 por ciento en la razón dos a uno propia de las desintegraciones puras de K_1. El error es suficientemente pequeño para excluir el cero. El experimento proseguirá durante otro año en un esfuerzo por comprender mejor todavía las fuentes de error.

Dentro de unos años los resultados de esa suerte de experimentos deberán determinar que la asimetría *CP* está causada por una fuerza milidébil o que la desviación de las desintegraciones puras es inferior al uno por ciento y que, por tanto, la asimetría *CP* probablemente esté causada por una fuerza superdébil. Cualquiera que sea el desenlace, impondrá restricciones relativas a la forma de la gran ecuación y eso nos acercará más a la comprensión de nuestro universo.



6. VIOLACION DE LA SIMETRIA *CP* por el mesón *K* neutro, de una manera distinta de la desintegración preferente en positrones (*véase la figura 4*). En la violación está involucrada la transición de un mesón K_2, que tiene una función de onda antisimétrica con respecto al espejo *CP*, a un mesón K_1, dotado éste de una función de onda simétrica. La violación es análoga a la interacción de luz con una disolución de dextrosa. La analogía empieza con un haz de mesones K^0 y un haz de luz polarizada circularmente a derechas. La componente K_1 de los mesones K^0 se desintegra, dejando la componente K_2; simultáneamente, un trozo de lámina polarizante impide el paso de la componente horizontal de la luz, permitiendo pasar sólo a la componente vertical. A medida que los mesones K_2 prosiguen su recorrido, tiene lugar una violación de *CP* que convierte algunos de ellos en mesones K_1 (la combinación es el mesón K_L); análogamente, una disolución de dextrosa provoca el giro de una pequeña componente de la luz del estado polarizado verticalmente al estado polarizado horizontalmente. En consecuencia, se generan mesones K_1 donde antes no había y se genera luz polarizada horizontalmente donde era inexistente. La "rotación" de los mesones K_2 antisimétricos respecto a *CP*, dando lugar a mesones K_1 simétricos bajo *CP*, constituye la única violación conocida de la simetría *CP*. (Esquemas de J. Kuhl.)



Riesgo volcánico

Las recientes erupciones del monte Santa Elena y el Nevado del Ruiz, de efectos tan dañinos, han impulsado el desarrollo de la vulcanología, sobre todo en lo relativo a la evaluación de los riesgos y a la vigilancia de los volcanes

Juan Carlos Carracedo

Los volcanes han adquirido en los últimos años un innegable protagonismo, no sólo en el ámbito restringido de las ciencias de la Tierra, sino también en el interés del público general. Débese ello a las catástrofes, humanas y materiales, que han provocado las erupciones del Santa Elena y el Nevado del Ruiz, entre otros.

No parece, sin embargo, que, desde una perspectiva geológica, e incluso histórica, pueda decirse que estemos ante un incremento de la actividad volcánica del planeta. Lo que sí ha crecido es la densidad de población en zonas volcánicas activas, con el consiguiente riesgo para sus vidas. Peligro que ha generado, en un movimiento de defensa, una mayor inversión en el campo de la investigación científica y técnica para evitar o mitigar al menos los efectos adversos de las erupciones.

Por marcar un hito de referencia, podría decirse que ese creciente protagonismo comenzó a las 8,30 de la mañana del 18 de mayo de 1980, cuando se produjo la erupción del volcán Santa Elena, dormido desde hacía 150 años. Aunque la estrecha vigilancia que los vulcanólogos estadounidenses mantenían sobre el mismo —alertados desde un año antes por el inicio de la actividad sísmica y varias explosiones freáticas, producidas por la presión del vapor generado al entrar el magma en contacto con el agua del subsuelo— evitó males mayores, la tremenda energía liberada a lo largo de las pocas horas que duró la erupción —equivalente a unas 27.000 bombas atómicas similares a las arrojadas sobre Hiroshima, con una cadencia de una bomba por segundo— ocasionó la muerte de 57 personas y pérdidas materiales cercanas a los 1000 millones de dólares.

El 13 de noviembre de 1985 entró en erupción otro volcán, también largo tiempo dormido: el Nevado del Ruiz, en el extremo norte de la cordillera andina colombiana. Esta erupción, por otra parte de energía relativamente baja, depositó material incandescente sobre el casquete glaciar que corona permanentemente el volcán, fundiéndolo en parte y desencadenando flujos de lodo que acabaron, en pocos minutos, con la vida de 25.000 personas y produjeron daños cifrados en más de 300 millones de dólares. Si la erupción hubiera sido más potente y hubiera fundido todo el casquete glaciar, la catástrofe habría alcanzado proporciones continentales. (De todas formas, el Nevado del Ruiz tiene el triste récord de ser la segunda erupción en lo que va de siglo por el número de víctimas —la más cruenta fue la del Monte Pelée en 1902, en la Martinica— y la tercera desde el año 1700.)

Tal vez sean las erupciones volcánicas uno de los fenómenos naturales más complejos y con mayor gama de variaciones, lo que dificulta sobremanera la previsión de su comportamiento. Un volcán puede tener una erupción con estilos y mecanismos muy diferentes de los manifestados en episodios anteriores e incluso presentar variaciones drásticas durante la erupción en cuestión. Se puede, ello no obstante, establecer algunas generalidades.

No es fácil comenzar por la más necesaria: ¿qué es un volcán? De manera provisional definámoslo como un proceso complicado que da lugar a la generación en profundidad de magma, mezcla pastosa de silicatos total o parcialmente fundidos y gases en disolución. El magma producido, que determinará en gran parte el tipo de volcán, ascenderá por mecanismos muy variados y complejos hasta la superficie a través de un conducto volcánico, cuyo orificio de expulsión es el cráter. Una vez en contacto con la atmósfera, se liberan los gases —a veces con tal energía que forman chorros o columnas eruptivas que pueden alcanzar varios kilómetros de altura— y las rocas fundidas, o lava. Esta última puede salir en forma continua, y tenemos las coladas, o fragmentada en trozos de diverso tamaño, y distinguimos las cenizas (los más finos), los lapilli (trozos del tamaño de granos de arena) y las bombas o bloques (de proporciones mayores). Al conjunto de los materiales volcánicos fragmentarios se le da el nombre de piroclastos, en directa alusión a su aspecto de pellas de fuego que es el significado literal de ese cultismo griego.

Ahondando un poco más, la teoría de la tectónica de placas nos proporciona un esquema global para la generación de magmas y de sus manifestaciones superficiales: los volcanes. Dicha teoría no explica sólo la génesis del vulcanismo, sino también los principales procesos que se desarrollan en la Tierra: formación de los continentes y los océanos, las montañas, las fosas oceánicas, los terremotos, etcétera. Actividades todas ellas que resultan de los movimientos relativos de las grandes placas en que se encuentra dividida la litosfera terrestre. Las placas litosféricas se separan a partir de un eje o borde divergente, se van aproximando, chocan y pueden acabar sumergiéndose unas debajo de otras en los bordes convergentes, denominados en este último

1. ASPECTO DEL CAUCE ALTO DEL RIO GUALI tras el paso de los flujos de lodo. La fusión de parte del casquete glaciar del Nevado del Ruiz por la caída de productos volcánicos incandescentes, en la erupción del 13 de noviembre de 1985, desencadenó torrentes de agua, mezclada con materiales de la propia erupción, vegetación y rocas arrancadas de las laderas del volcán. Estos flujos de lodo o lahares, de elevada densidad y volumen —más de 60 millones de metros cúbicos— se precipitaron a gran velocidad por la red de barrancos, destruyendo la red viaria local (doscientos kilómetros de carretera y 19 de puentes, por lo menos), aislando comunidades enteras y arrasando parcialmente el pueblo de Chinchiná y, totalmente, el de Armero.

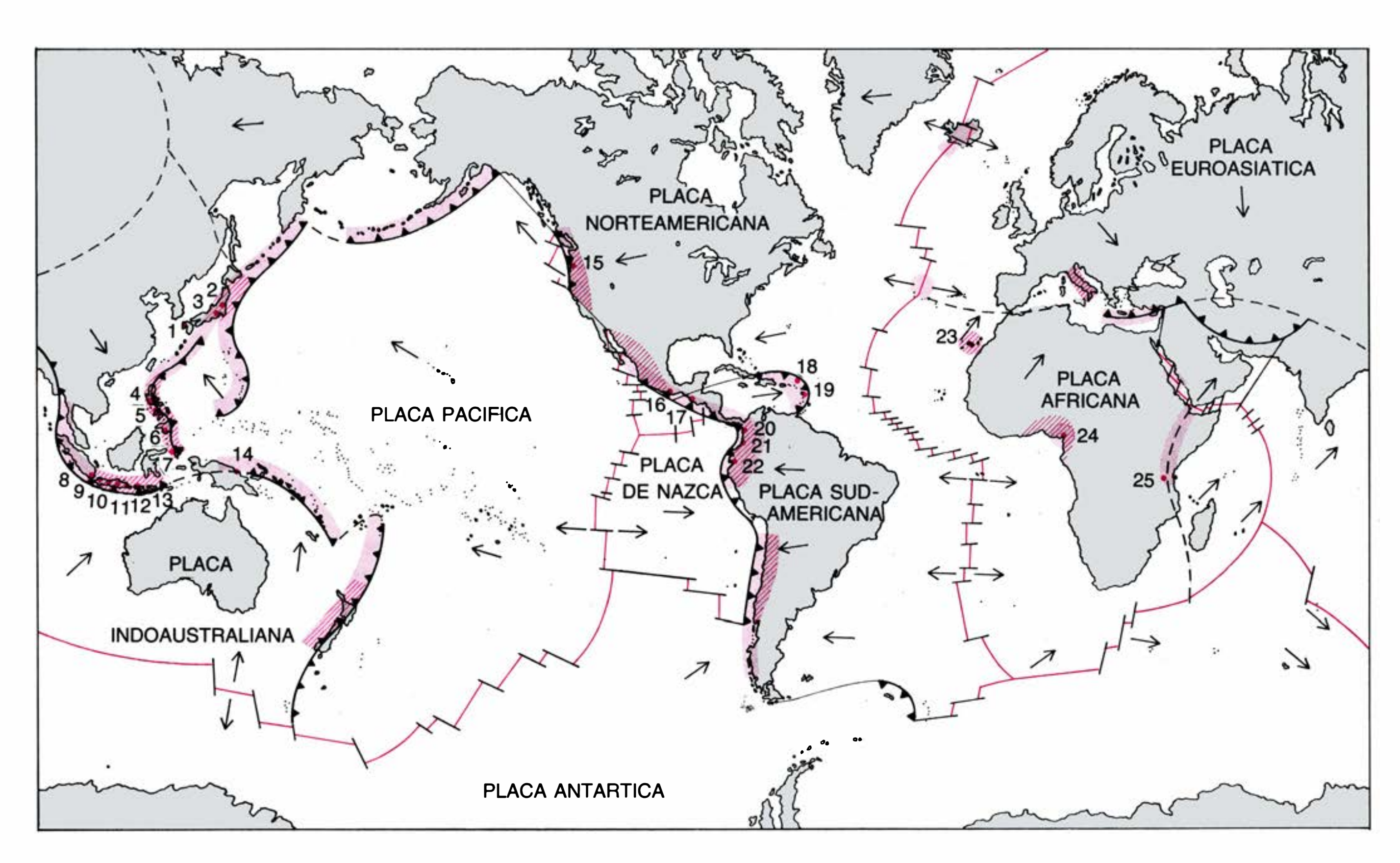
caso "activos" por estar asociados a una intensa actividad sísmica y volcánica.

La mayoría de los volcanes, el 95 por ciento, se sitúan en los bordes de esas placas litosféricas y sólo el 5 por ciento restante lo hace en su interior. En términos generales, tanto el vulcanismo de borde divergente como el de interior de placa suelen generarse a partir de magmas poco viscosos, con mecanismos eruptivos de baja explosividad. El riesgo de ambos tipos de erupción va asociado, fundamentalmente, a la caída de piroclastos (lapilli y bombas) en un entorno restringido del centro de emisión y al flujo de lavas muy fluidas, que a veces alcanzan grandes velocidades. El daño a las personas suele ser mínimo o nulo y afecta sólo a las estructuras que no pueden desplazarse: carreteras, puertos, viviendas, flora, etcétera.

Un caso excepcional lo constituyó la erupción del volcán zaireño Nyiragongo, en 1977. Un lago de lavas muy fluidas que ocupaba el cráter de 1200 metros de diámetro y 1000 de profundidad se vació rápidamente a través de un sistema de fisuras paralelas que se abrieron simultáneamente en los flancos del volcán. En poco más de una hora, diez millones de metros cúbicos de lava fluyeron ladera abajo a velocidades de hasta 60 kilómetros por hora, con frentes de dos metros de altura. El balance final de ese insólito fenómeno volcánico señalaba la pérdida de cinco kilómetros cuadrados de cultivo, centenares de casas destruidas y de 50 a 100 víctimas humanas.

En los bordes convergentes, en cambio, los magmas generados suelen ser muy viscosos. Atrapan el gas disuelto y alcanzan altísimas presiones que se liberan, de forma explosiva, en la erupción. A esta última situación se puede llegar también partiendo de magmas fluidos, a través de procesos que cambian la composición inicial del magma aumentando la viscosidad. Pertenecen a este tipo de magmas viscosos la mayoría de los grandes volcanes del planeta (el Cinturón de Fuego del Pacífico, con el 75 por ciento del vulcanismo activo) y, desde luego, los más peligrosos. Característica de esos volcanes es la permanencia de la actividad eruptiva concentrada en un punto durante largo tiempo; en ellos también, la acumulación de materiales de sucesivas erupciones y el escaso recorrido de las lavas a causa de su viscosidad, crea edificios de gran altura. Y en esa progresiva elevación de los edificios volcánicos se encierra el germen de su propia destrucción, aumentando correlativamente su peligrosidad. En efecto, el aumento en altura convierte a los edificios en mecánimente inestables, facilita su colapso y desencadena, a veces, erupciones catastróficas, al liberar de forma instantánea la presión acumulada en el interior.

En el comienzo de estas erupciones suele ocurrir que el conducto volcánico esté taponado por lava solidificada de las emisiones anteriores; la presión de los gases encerrados puede así adquirir valores altísimos, en particular si el

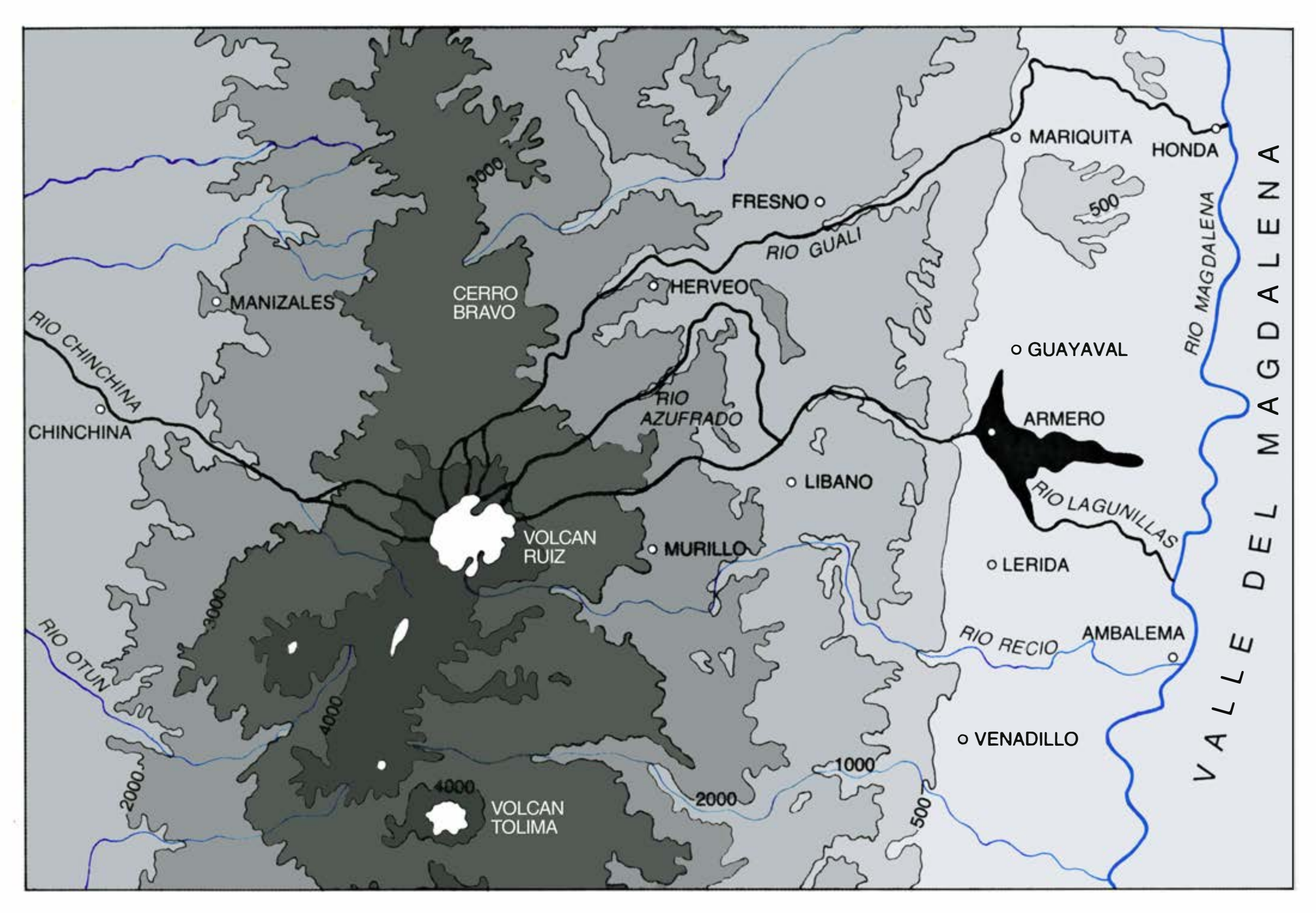


• VOLCANES : 1 UNZEN 2 ASAMA 3 OSHIMA 4 TAAL 5 MAYON 6 AWU 7 MAKIAN 8 KRAKATAU 9 PAPANDAJAN 10 MERAPI 11 KELUT 12 AGUNG 13 TAMBORA 14 LAMINGTON 15 ST. HELENS 16 CIHCHON 17 STA. MARIA 18 MT. PELEE 19 SOUFRIERE 20 NEVADO DEL RUIZ 21 TOLIMA 22 COTOPAXI 23 TEIDE 24 MT. CAMEROUN 25 NYIRAGONGO

AREAS DE INTENSO VOLCANISMO — AREAS VOLCANICAS ACTIVAS CON ALTA DENSIDAD DE POBLACION

BORDE DE PLACA CONVERGENTE — BORDE DE PLACA DIVERGENTE — ⟵ SENTIDO DE DESPLAZAMIENTO DE LAS PLACAS

2. UBICACION EN EL MARCO DE LA TECTONICA DE PLACAS de los volcanes que han producido las erupciones con mayor número de víctimas desde el año 1700. El riesgo volcánico en el planeta es elevado, al coincidir algunas de las zonas más densamente pobladas con las áreas de vulcanismo activo. Este riesgo es agudo en los bordes convergentes de placas —el Cinturón de Fuego del Pacífico—, donde los mecanismos eruptivos poseen un carácter violento.



3. COMPLEJO VOLCANICO RUIZ-TOLIMA, situado en el extremo norte de los Andes colombianos, formando una cresta entre los valles de los ríos Cauca y Magdalena. El gran desnivel existente entre la cima del Ruiz, a 5400 metros, y el valle del Magdalena, a poco más de 400 metros, salvado en tan sólo 40 kilómetros de recorrido, explica la elevada energía adquirida por los flujos de lodo originados al fundirse parte del casquete glaciar (*color blanco*) en la erupción del cráter Arenas (*punto rojo*). La canalización de estos flujos de lodo (*trazo negro*) por la red de barrancos llevó su poder destructor hasta las poblaciones.

magma entra, en su ascenso, en interacción con el agua del subsuelo, pues se producen entonces grandes cantidades de vapor de ésta a gran presión. El equilibrio entre la presión confinante de las rocas y la fuerza expansiva del gas puede romperse, bien porque esta última venza la resistencia mecánica del tapón que obstruye el conducto de salida, bien por un fenómeno añadido: un movimiento sísmico intenso que provoque el colapso del edificio inestable. En ambos casos, la rotura es análoga a la producida por un explosivo, originándose una onda expansiva dirigida, a veces lateralmente, en cuyo caso adquiere enorme poder destructivo. En la erupción del santa Elena se había producido una protuberancia por efecto de una intrusión lateral de magma, que inestabilizó el flanco norte del edificio y obturó la salida de los gases y del vapor de agua, mantenidos así a enorme presión. Un terremoto de magnitud cinco precipitó el colapso del flanco entero del volcán y provocó una explosión lateral dirigida, cuya onda de choque barrió 500 kilómetros cuadrados a velocidades de 400 kilómetros por hora, que arrancó de cuajo árboles de hasta dos metros de diámetro. Simultáneamente, una avalancha de dos kilómetros cúbicos de rocas y hielo, fluidificada por el vapor de agua, se precipitó ladera abajo, alcanzando velocidades de 250 kilómetros por hora, que le permitían remontar pendientes de hasta 400 metros.

Tanto las explosiones laterales como las avalanchas producen la destrucción total del área afectada, y no hay otra forma de evitar sus efectos que la evacuación previa de la zona.

Una vez que el conducto eruptivo se pone en comunicación directa con la atmósfera, la brusca caída de presión libera el gas ocluido a sobrepresión en el magma viscoso, generándose burbujas que acaban estallando y fragmentando el magma. Se pasa así súbitamente de un sistema líquido-gas a otro disperso de gas-lava fragmentada, lo que conlleva un descenso general de la presión en el conducto de salida y una enorme expansión del segundo sistema.

El proceso en cadena que se genera da lugar a un chorro o columna eruptiva que se dispara verticalmente a velocidades de hasta 600 metros por segundo y en continua expansión; arroja a la atmósfera gases sobrecalentados, fragmentos de lava incandescente y trozos arrancados de las paredes del conjunto en volúmenes que a veces se miden por kilómetros cúbicos (1,4 en el episodio del santa Elena). La expansión, que prosigue al inyectarse el chorro en la atmósfera, provoca una pérdida de densidad por caída de los fragmentos menos ligeros, lo que eleva la columna hasta alcanzar frecuentemente decenas de kilómetros de altura. Al llegar a la zona de inversión de temperaturas (límite estratosfera-troposfera) adquiere la forma de hongo típica y, por último, arrastrada por los fuertes vientos laterales, la de pluma.

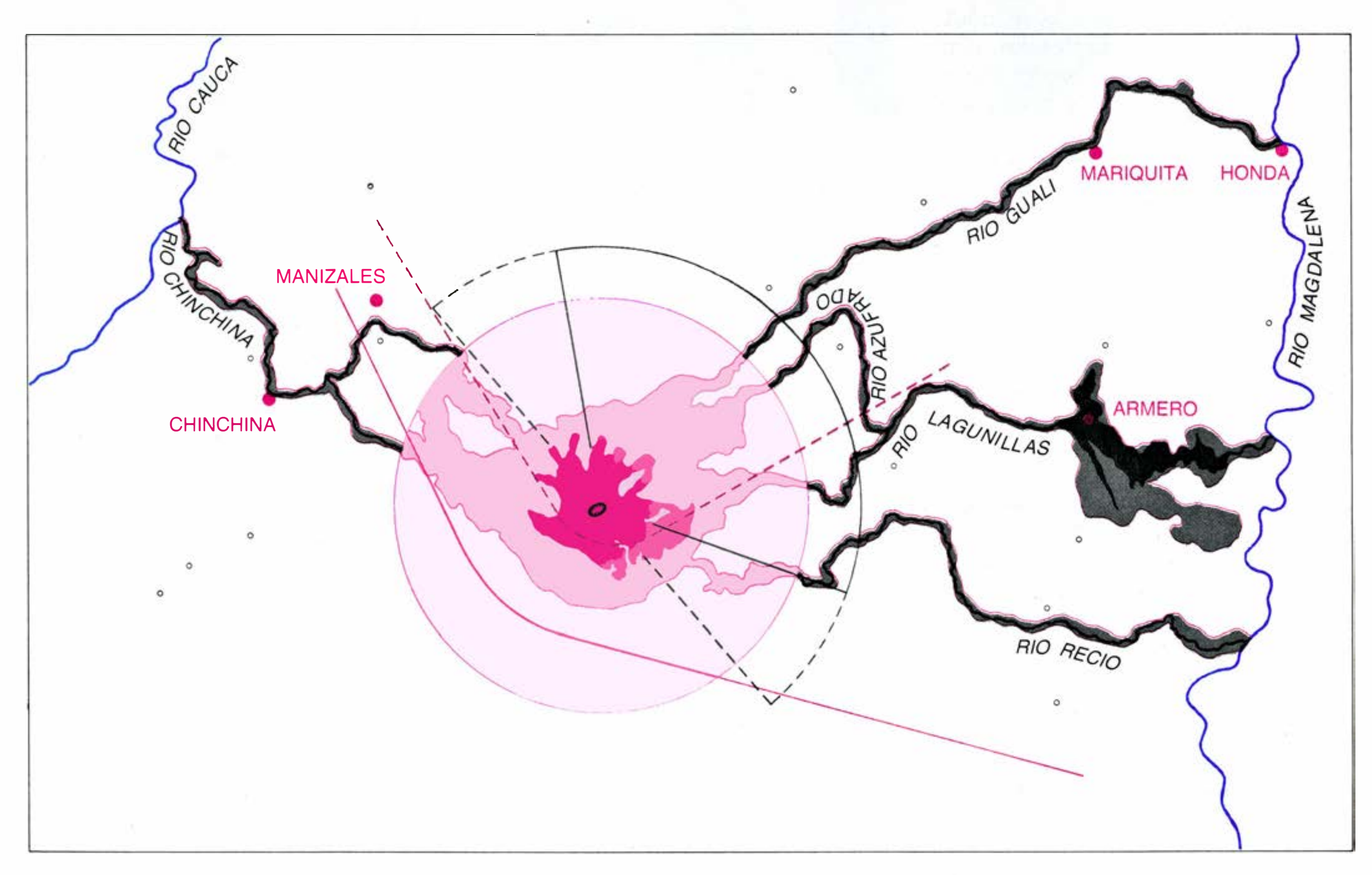
Según la altura que alcancen el chorro o la pluma, y de acuerdo con la dirección y velocidad del viento, los fragmentos se dispersan por una amplia zona. Los de granulometría muy fina —las llamadas cenizas— se mantienen

largo tiempo en suspensión y llegan hasta puntos muy alejados del epicentro del episodio. En la erupción del santa Elena se observaron depósitos de ceniza de varios centímetros de espesor en varias ciudades de Washington, Idaho y Montana; en la del Nevado del Ruiz se registró caída de cenizas en Venezuela, a 400 kilómetros de distancia.

A diferencia de las explosiones laterales dirigidas y las avalanchas, las lluvias de cenizas no representan un grave riesgo para las personas y la red viaria de una zona. Lo que no deja de ser una suerte, pues se trata del fenómeno más típico de las erupciones volcánicas. Producen, sin embargo, graves trastornos en los núcleos de población: hundimiento de techos de viviendas, intensa reducción de la visibilidad, daños en motores de combustión, alteración del suministro energético, etcétera. Acarrean también graves perjuicios contra los recursos agropecuarios, al destruir cosechas y provocar -por ingestión de las cenizas volcánicas- trastornos digestivos e incluso la muerte del ganado.

Las coladas o flujos de piroclastos son, sin duda, el fenómeno más cruento y destructivo de estas erupciones volcánicas. Su génesis está asociada al ascenso del chorro eruptivo a la atmósfera, en el que se distinguen dos fases: una superior convectiva —que dará lugar a la lluvia de cenizas— y otra inferior de empuje, en que la gravedad frena rápidamente el ascenso de los fragmentos sólidos (piroclastos) por efecto de la velocidad inicial y provoca el colapso de los mismos y su rápido descenso por las laderas del volcán. Las temperaturas de centenares de grados de estos flujos piroclásticos —llamados, con toda propiedad, nubes ardientes—, sus velocidades a veces de cientos de kilómetros por hora y su enorme movilidad, favorecida por la fluidificación del sistema (descenso de la densidad y viscosidad) por efecto de los gases calientes que mantienen en flotación las partículas sólidas, explican los letales efectos del fenómeno. Los edificios, los bosques y cualquier estructura en pie resultan arrasados y quemados hasta su total destrucción; las probabilidades de supervivencia bajo el impacto directo de una colada piroclástica son virtualmente nulas, sin que puedan tomarse otras medidas que la evacuación a tiempo de las zonas amenazadas.

En este breve repaso de los principales factores de riesgo asociados a las erupciones volcánicas queda por citar uno de los más frecuentes y peligrosos: los flujos de lodo o lahares. El crecimiento en altura de los edificios volcánicos poligénicos —llamados así por haberse construido por acumulación de materiales de varias erupciones, en contraposición a los monogénicos que son el resultado de una sola erupción— no sólo puede provocar su debilitamiento y colapso, sino que genera laderas con pendientes muy pronunciadas y grandes desniveles, proporcionando una elevada energía potencial a los materiales que se acumulan en precario equilibrio en la cima del edificio. Basta a veces una ligera lluvia para desencadenar un flujo de lodo



LAHAR PREVISTO — LAHAR PRODUCIDO — RIESGO ALTO DE EXPLOSION LATERAL DIRIGIDA — RIESGO MODERADO DE EXPLOSION LATERAL DIRIGIDA — RIESGO ALTO DE CAIDA DE PIROCLASTOS — RIESGO MODERADO DE LLUVIA DE PIROCLASTOS — RIESGO ALTO DE FLUJO DE COLADAS LAVICAS — RIESGO MODERADO DE FLUJO DE COLADAS LAVICAS — RIESGO ALTO DE FLUJOS PIROCLASTICOS — RIESGO MODERADO DE FLUJOS PIROCLASTICOS

4. MAPA DE RIESGOS VOLCANICOS DEL NEVADO DEL RUIZ, elaborado por Ingeominas y presentado una semana antes de la erupción del 13 de noviembre de 1985. Difundido por Colombia a través de un periódico nacional, este trabajo, que anticipa con gran precisión las zonas que luego serían devastadas por los flujos de lodo, pudo ser un elemento decisivo en la previsión y mitigación de los efectos de la erupción. De los riesgos indicados en el mapa, sólo el de los lahares se produjo de forma muy similar a la prevista. No ocurrió así con el flujo de coladas lávicas, que no se dio, ni con las explosiones laterales.

fluidificado, que, a causa de su densidad y de las fuertes pendientes, se canaliza por los barrancos y avanza a velocidad de hasta decenas de kilómetros por hora. En su recorrido recogen material arrancado de los cauces, aumentando así su densidad y su poder destructor. Por otro lado, los volcanes de gran altura suelen estar cubiertos por casquetes permanentes de hielo; la presencia de millones de metros cúbicos de agua, a miles de metros de altura, supone un riesgo terrible, pues en caso de erupción la caída de material a elevada temperatura provoca la rápida licuación del hielo y genera flujos de lodo devastadores.

Un ejemplo desgraciadamente espectacular del poder destructor de los flujos de lodo fue la erupción del 13 de noviembre de 1985 del Nevado del Ruiz. Este volcán gigantesco, que se encumbra a 5400 metros de altura, se halla en el extremo septentrional del complejo volcánico Ruiz-Tolima, en los Andes colombianos. Pertenece al tipo de vulcanismo propio de margen activo, donde los magmas suelen ser viscosos y acostumbran a darse mecanismos eruptivos muy explosivos. Una densa red de barrancos radiales une el casquete glaciar, de 21 kilómetros cuadrados y unos 500 millones de metros cúbicos de hielo que corona este volcán, con los valles de Cauca y el Magdalena. Los ríos Lagunillas y Azufrado descienden desde la cima del Nevado del Ruiz hasta el pueblo de Armero, salvando en poco más de 40 kilómetros un desnivel de 4500 metros.



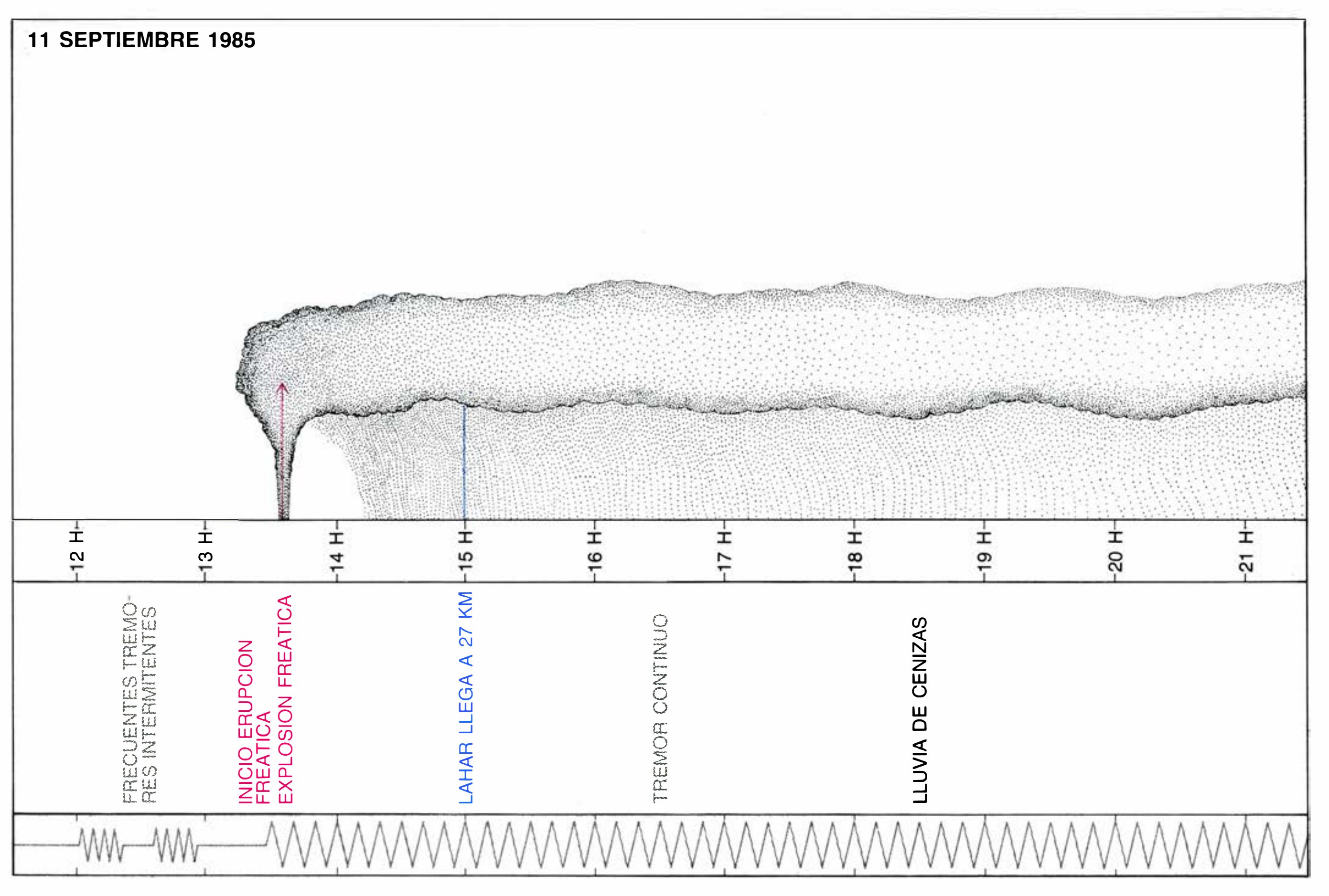
5. **LA EXPLOSION FREATICA producida a las 15,05 horas, al inicio de la erupción del 13 de noviembre, depositó un fino tapiz de cenizas volcánicas sobre la vegetación en un radio de decenas de kilómetros. La ganadería sufrió graves trastornos al ingerir estos productos, tóxicos y con fragmentos cortantes de vidrio volcánico. La muerte de numerosas reses obligó a desplazamientos masivos de ganado a zonas alejadas.**

La erupción del 13 de noviembre de 1985, aunque de escasa duración y baja energía, produjo la caída de fragmentos de pómez incandescentes y flujos piroclásticos sobre el casquete glaciar. La brusca fusión de unos 50 millones de metros cúbicos de agua provocó lahares que transportaron por los cauces de estos ríos entre 30 y 40 millones de metros cúbicos de lodo a velocidades en ocasiones superiores a los 40 kilómetros por hora, hasta una extensión de 80 kilómetros. En su recorrido asolaron la localidad de Armero, pereciendo en pocos instantes 22.000 habitantes. La magnitud de la catástrofe se pondera mejor si se considera que sólo se fundió apenas un 10 por ciento del volumen total del glaciar. De haber alcanzado la erupción mayor intensidad y se hubiese fundido todo el glaciar, las proporciones del siniestro hubieran sido inimaginables.

Pero el desastre de Armero pudo haberse evitado. ¿Cómo? Aunque este volcán llevaba largo tiempo inactivo y se había convertido incluso en centro de atracción turística, los geólogos conocían su historia, que registraba al menos 10 grandes erupciones de carácter explosivo a lo largo de los últimos 10.000 años. En 1595 y 1845 se produjeron erupciones, siendo el último episodio responsable de flujos de lodo que barrieron el cauce del río Lagunillas hasta alcanzar el Magdalena. Armero quedó arrasado y murieron mil personas.

Un año antes de la catastrófica erupción de 1985, el Nevado del Ruiz dio claros indicios de reanudar su actividad, con un incremento anormal de su sismicidad y producción fumaroliana. Expertos colombianos y extranjeros avisaron a las autoridades de una previsible erupción con grandes probabilidades de originar flujos de lodo destructivos. Se inició la elaboración de un mapa de riesgos volcánicos del Ruiz y se instalaron cuatro sismógrafos portátiles alrededor del volcán, que confirmaron inmediatamente el elevado nivel de actividad sísmica. El 11 de septiembre se produjo una erupción freática; a 25 kilómetros del volcán, sobre la ciudad de Manizales (230.000 habi-

NOVIEMBRE
DICIEMBRE
ENERO
FEBRERO
MARZO
ABRIL
MAY
1984
1985
DIA 22
FRECUENTES TERRE-
MOTOS DE BAJA
INTENSIDAD
MAGNITUD 3-4
EXPLOSION FREATICA
LLUVIA DE LITICOS
FUMAROLA
INCREMENTO
FUMAROLA
FUMAROLA



6. SECUENCIA DE LAS DIFERENTES FASES de actividad del volcán Nevado del Ruiz, desde los primeros indicios de reactivación a finales de 1984 hasta la erupción pliniana, catastrófica, del 13 de noviembre de 1985. Los fenómenos sísmicos se indican de forma muy simplificada en las casillas horizontales y con

MAYO | JUNIO | JULIO | AGOSTO | SEPTIEMBRE | OCTUBRE | NOVIEMBRE

DIA 20
INSTALACION SISMO-GRAFOS PORTATILES

DIA 11
TREMOR INTERMI-TENTE
ERUPCION FREATICA
EXPLOSION FREATICA

DIA 15
LAHAR (27 KM)
LLUVIA DE LITICOS

DIA 7
PRESENTACION Y PU BLICACION DEL MAPA DE RIESGOS VOLCA-NICOS
INSTALACION INCLINOMETROS
DESCENSO ACTIVIDAD
TREMOR VOLCANICO CONTINUO

DIA 13
ERUPCION PLINIANA
EXPLOSION FREATICA
LAHARES + 25.000
COLUMNA ERUPTIVA PLINIANA (10 KM)
CAIDA PIROCLASTOS INCANDESCENTES

13 NOVIEMBRE 1985

15 H
SATURACION DE REGISTRO
EXPLOSION FREATICA

16 H
LLUVIA DE CENIZAS EN MARIQUITA

17 H
LLUVIA DE CENIZAS EN ARMERO

18 H

19 H
SE ORDENA EVACUA-CION DE ARMERO. NO SE REALIZA

20 H
FLUJOS PIROCLASTI-COS

21 H
EXPLOSIONES FREATI-CAS DIRIGIDAS
SATURACION SISMO-GRAFOS
COMIENZA LA FUSION DEL HIELO

22 H
AVISO POR RADIO A ARMERO
PIROCLASTOS INCANDESCENTES
INTENSA LLUVIA CENI-ZAS SOBRE ARMERO
LAHAR ALCANZA CHINCHINA + 1.000

23 H
LAHAR ALCANZA ARMERO + 22.000
PLUMA EN PROGRE SION CRECIENTE: 10 KM
RUIDO SISMICO PRO-DUCIDO POR LAHARES

24 H

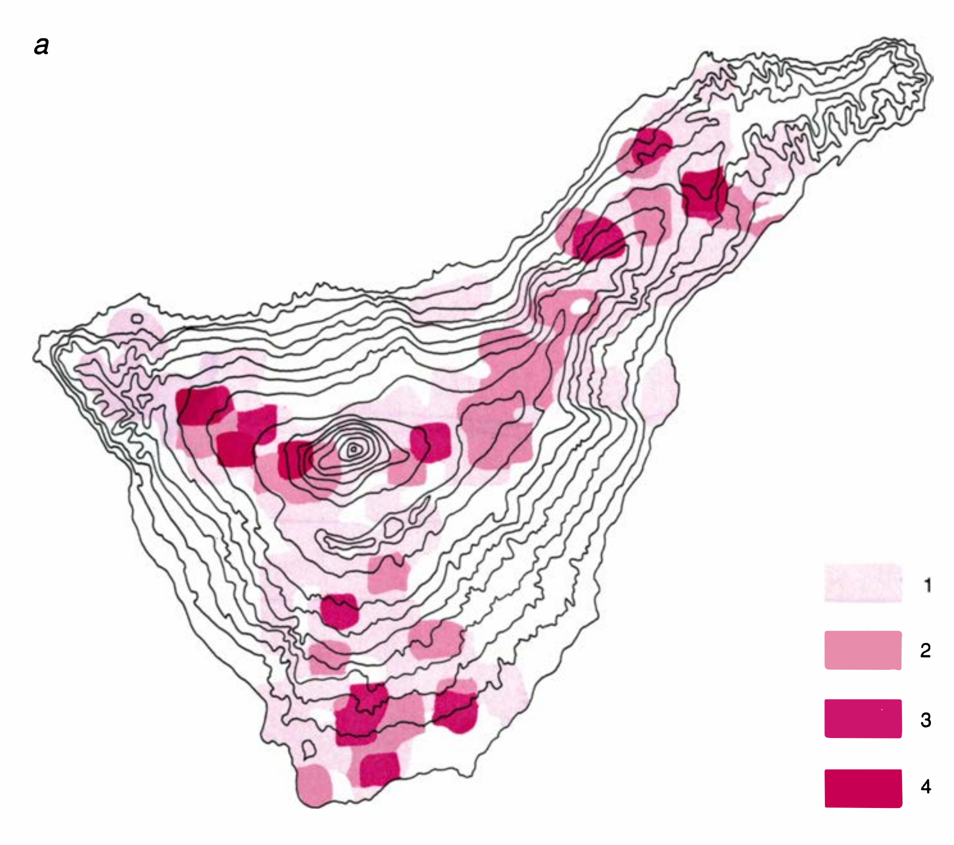
rotulación de color gris; los episodios eruptivos en rojo y en azul señalan los hitos principales relacionados con los flujos de lodo. El tiempo se indica en hora local. (El esquema se ha elaborado con datos del Comité de Estudios Volcanológicos del Nevado del Ruiz, del que formaba parte el autor de este artículo.)

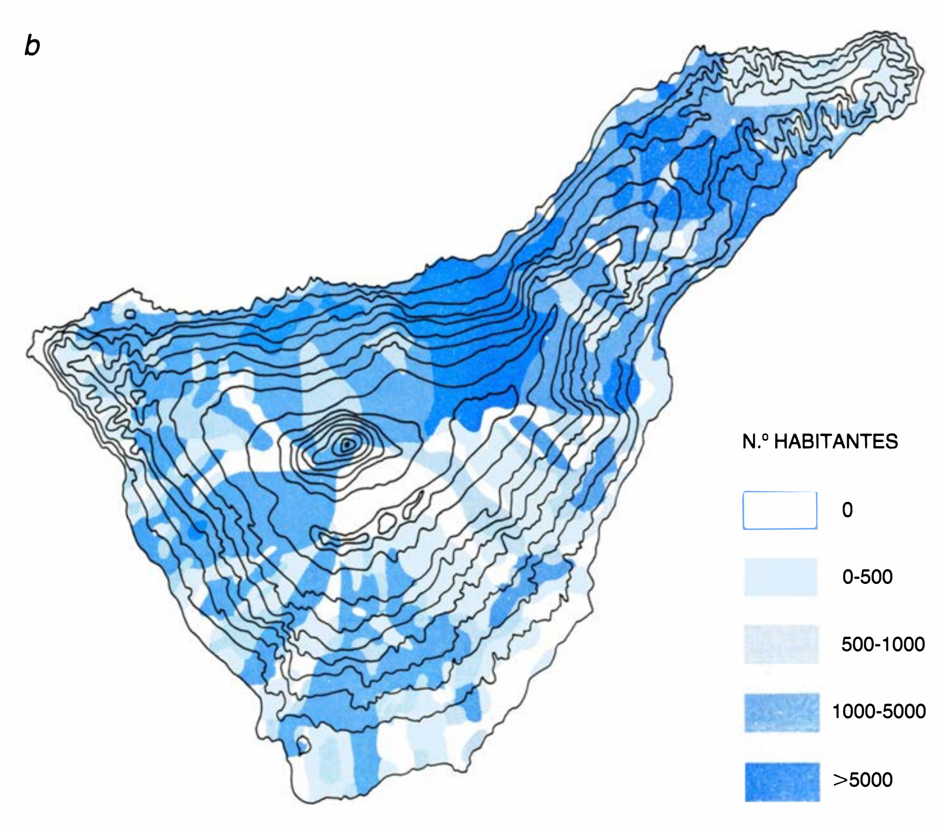
tantes), cayó una lluvia de cenizas y comenzó a fluir lodo por los ríos Guali y Azufrado, que progresó en su avance hasta 27 kilómetros de distancia. Resultaba obvio que se vivía en una situación de alto riesgo de erupción importante.

El mapa de riesgos volcánicos se terminó de elaborar el 7 de octubre y en él se anticipaba con gran exactitud las zonas que serían luego devastadas por los lahares, Armero incluido. Se entregó el mapa a las autoridades y se publicó en la primera página de un periódico nacional; se exponía en él con claridad que, de producirse una erupción, habría un 100 por ciento de probabilidades de que se originasen flujos de lodo que podían asolar las poblaciones de Chinchiná, Honda, Mariquita y, especialmente, la de Armero. Se recomendaban también medidas de evacuación. En particular, se aconseja la evacuación inmediata a zonas altas al observar caída de cenizas, indicio claro del comienzo de la erupción y de la inminente llegada de los flujos de lodo. Pero la escasa percepción de la amenaza por parte de las autoridades y la población, así como un conjunto de hechos desgraciados, permiten que ocurra una catástrofe evitable.

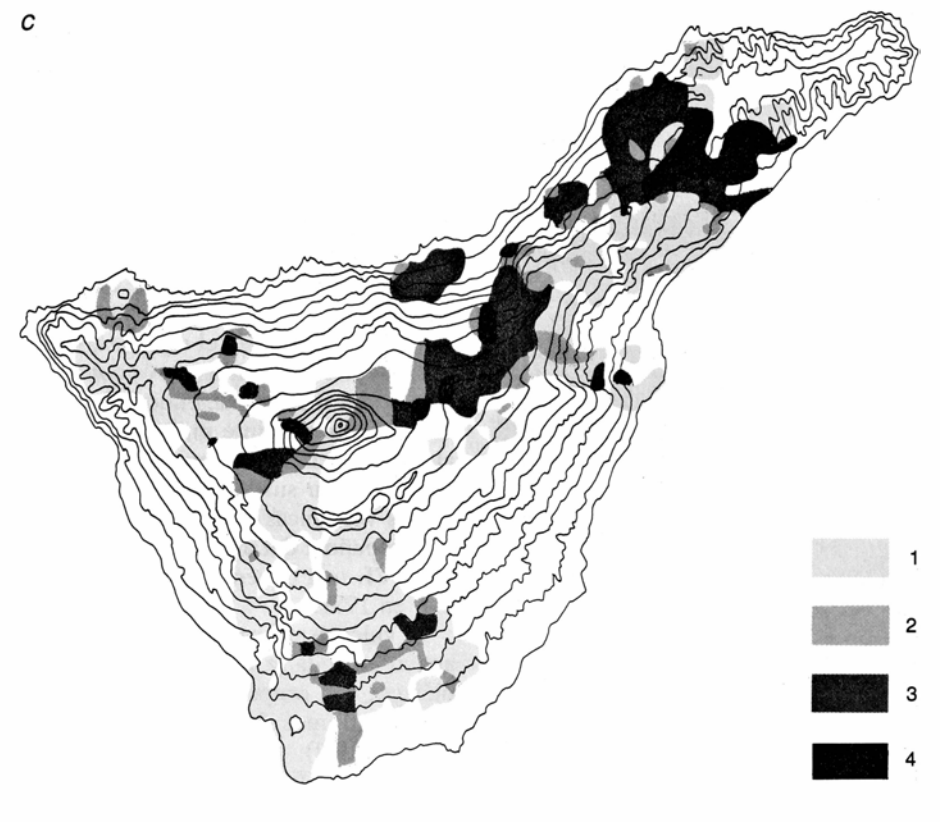
Aunque no se percibieron signos previos especialmente alarmantes —el mismo día precedente a la erupción varios geólogos visitaron el interior del cráter—, ésta se inició de repente el 13 de noviembre, poco después de las tres de la tarde. Una fuerte explosión freática dispersó finos fragmentos de pómez y rocas en un radio de 50 kilómetros, al tiempo que se formó una densa columna negra en la cima del volcán. Dos horas más tarde, comenzaba a caer una fina lluvia de cenizas sobre Armero. Avisadas las autoridades, se decidió finalmente, a las 19,30 horas, la evacuación de Mariquita y Armero, evacuación que no se llevó a efecto tal vez por la resistencia de la población a interrumpir la cosecha de café y el temor a saqueos. Poco después de las nueve de la noche comenzaba la fase paroxística de la erupción: dos violentas explosiones freáticas que se oyeron a 30 kilómetros del volcán coincidieron con el inicio de una vibración continua o tremor armónico que saturó los registros de los sismógrafos (21,05 hora local) por espacio de casi una hora.

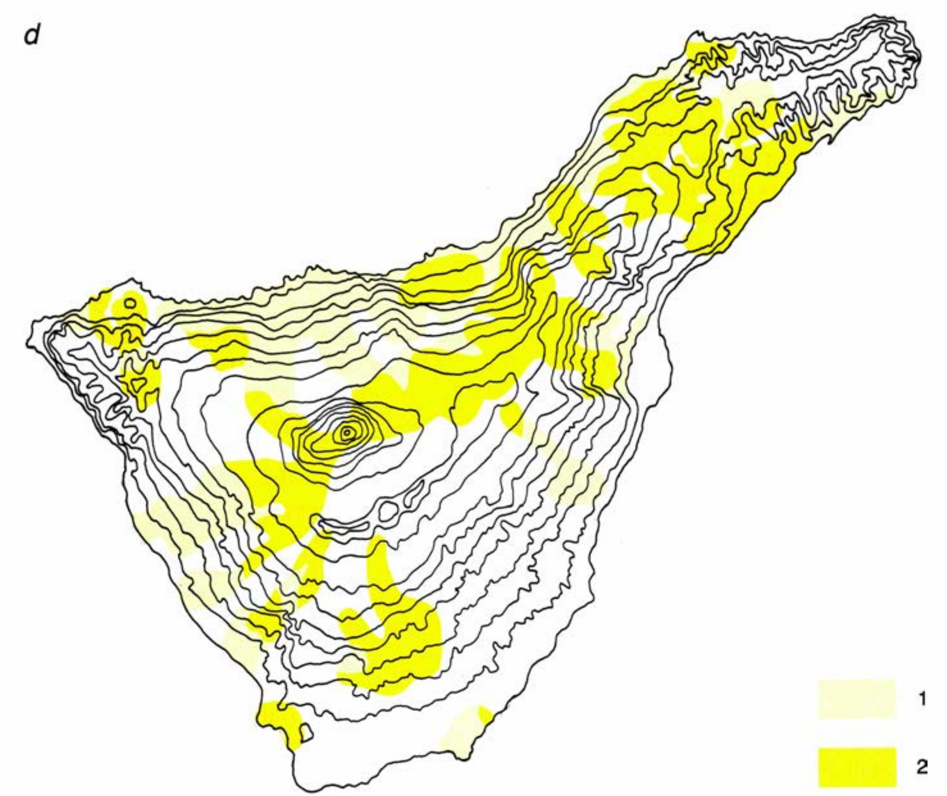
Violentas explosiones freáticas laterales de carácter turbulento y flujos piroclásticos destruyeron e incendiaron el motel refugio situado a dos kilóme-





7. MAPAS DE RIESGO VOLCANICO de las dorsales activas de la isla de Tenerife. Para su elaboración, realizada por ordenador, se dividió la isla en cuadrículas de un kilómetro cuadrado, asignando a cada una de ellas 17 parámetros relacionados con la población, infraestructura, datos vulcanológicos y pendiente topográfica. La simulación de erupciones se basa en la estimación de la dispersión de piroclastos y en el flujo de coladas siguiendo la máxima pendiente. La interacción de varios archivos de datos —densidad de población e infraestructura, concentración de centros de emisión recientes y topografía— permite, mediante pro-





cesado, la obtención de varios tipos de mapas de riesgo. El mapa *a* delimita zonas con distintos niveles de probabilidad de ocurrencia de erupciones (1 bajo, 2 moderado, 3 alto y 4 muy alto), en función del número de centros de emisión recientes. El mapa *b* indica la población que quedaría afectada por una erupción en cualquier punto de la isla, sin introducir el factor probabilidad. El mapa *c* incorpora ese factor, definiendo así zonas de distinto nivel de riesgo. Por último, el mapa *d* indica las zonas donde una erupción afectaría a núcleos de población de más de 1000 habitantes, con baja (*1*) y elevada (*2*) probabilidad de ocurrencia.

tros al oeste del cráter, depositando una fina capa de lodo en los obstáculos verticales y arrancando la vegetación hasta un radio de cinco kilómetros en algunos puntos. A las 21,30 comenzó a formarse una columna eruptiva que llegaría, en progresión creciente, hasta una altura de más de 10 kilómetros. De la columna se desprendieron fragmentos de roca arrancadas del conducto de salida y escorias y pómez incandescentes, que fueron a caer sobre el glaciar durante al menos hora y cuarto.

Los flujos piroclásticos y la lluvia de fragmentos de lava incandescente produjeron la rápida licuación de grandes volúmenes de hielo que se convirtieron en torrentes de agua (21,20 horas), que arrastraron consigo buena parte de los piroclastos depositados sobre el glaciar y materiales arrancados de las laderas, transformándose en flujos de lodo que descendieron por los flancos del volcán y terminaron canalizados por la red de barrancos. La relativa escasa importancia de esta erupción, tanto por su duración como por el volumen de materiales emitidos (de 10 a 100 millones de metros cúbicos, según estimaciones sobre el terreno), demostraba la eficaz transferencia de calor del material fragmentario al hielo, si tenemos en cuenta el volumen de hielo fundido. La generación de los lahares se vio favorecida por la acumulación de materiales sueltos en la base del glaciar y por las fuertes lluvias que precedieron a la erupción, que los transformaron en una masa saturada de gran inestabilidad.

Alrededor de las 22 horas, se avisó por radio a la ciudad de Armero, pero, al parecer, en vez de ordenarse la evacuación se aconsejó a sus habitantes permanecer en sus casas. El frente de flujo de lodo que discurría por el río Molinos llegó , a las 22,40, al pueblo de Chinchiná: destruyó 400 casas y hubo mil bajas. La lluvia de cenizas se hizo más intensa sobre Armero y el frente llegaba y arrasaba la población a las 23, 35. Todavía tenían tiempo las personas que permanecían despiertas de huir, alertadas por el fragor del flujo que se avecinaba con un frente de varias decenas de metros de altura. De una caravana de unos 50 coches, los 10 de cabeza consiguen ponerse a salvo, prueba de que, si se hubiera preparado a la población al comienzo del episodio, se habría contado con tiempo suficiente para su total evacuación. (Conviene no olvidar, sin embargo, que toda evacuación implica siempre cuantiosos perjui-



8. NEVADO DEL RUIZ, unas semanas después de la erupción del 13 de noviembre de 1985, visto desde la ciudad de Manizales. La espectacular columna que se eleva del cráter está compuesta principalmente por vapor de agua. La proximidad del volcán explica la aguda percepción actual del riesgo por parte de la población, en contraste con la época anterior a la erupción, en que se había perdido memoria de su peligrosidad por su inactividad desde el año 1895, hasta el punto de haberse convertido en un importante recurso turístico.

cios económicos y graves trastornos para los individuos afectados. Resulta muy difícil, y subjetivo, ponderar cuándo se debe decidir la evacuación, sobre todo si no hay antecedents en la zona o la población no está bien informada.)

Desgraciadamente han tenido que ocurrir esas catástrofes para que la vulcanología conociera un espectacular desarrollo en lo concerniente a la evaluación del riesgo volcánico y la vigilancia y previsión de las erupciones volcánicas. El comportamiento aleatorio de los volcanes impide, sin embargo, una predicción precisa de la ocurrencia de un episodio y del estilo y la magnitud del mismo.

La defensa contra sus efectos descansa, pues, en la elaboración de mapas que determinan los riesgos previsibles, mediante la reconstrucción de la historia del volcán, en la vigilancia con técnicas que permitan la detección con antelación del inicio de la erupción y la planificación anticipada de las medidas a adoptar al producirse la crisis. En este orden resulta ejemplar la vigilancia que ejerce del Nevado del Ruiz el Comité de Estudios Volcanológicos; en este volcán, todavía activo y con parecido nivel de riesgo que antes de la erupción del 13 de noviembre, se mantiene bajo observación cualquier variación en su forma o en la del cráter, aparición o ensanchamiento de fisuras en el volcán y en el casquete glaciar, incremento de la actividad fumaroliana, emisión de materiales sólidos, estudio de la red de drenaje y otros. Súmese a ello la observación directa periódica, la fotografía aérea y el registro sísmico y de otros tipos.

Afortunadamente, la erupción no es el inicio sino la fase final del verdadero fenómeno volcánico, que consiste en la generación de magma y su ascenso a la superficie. El violento proceso de desplazamiento de una masa de gran volumen a muy elevada temperatura produce unos efectos físicos y químicos que pueden detectarse, actuando como precursores de la erupción. El Nevado del Ruiz cuenta hoy con una red de seis sismógrafos colocados en un radio de unos cinco kilómetros alrededor del cráter, que transmiten la señal en tiempo real por radio al Observatorio en Manizales; esta red permite el registro continuo de los seísmos y la localización de sus focos, aportando información sobre cualquier incremento de la actividad del volcán y sobre posibles cambios en el nivel de magma. Otra red de cuatro inclinómetros para registrar hasta las mínimas variaciones de nivel informan en tiempo real de cualquier deformación del terreno, detectando procesos de inflación-deflación originados por ascensos del magma desde zonas más profundas.

En España, el riesgo volcánico se circunscribe a las islas Canarias. Aunque hay vulcanismo peninsular relativamente reciente, desde el punto de vista geológico, únicamente el de Canarias puede considerarse activo. El propio proceso volcánico es el responsable de la existencia de las islas, cuya construcción se ha ido produciendo a lo largo de los últimos veinte millones de años al menos. En los últimos milenios, las Canarias han sufrido cientos de erupciones volcánicas y, desde la colonización española del archipiélago en el siglo XVI, una docena de ellas. La última, en 1971.

El vulcanismo canario se manifiesta activo en toda la alineación del archipiélago, con erupciones muy recientes en sus extremos: Lanzarote (1824) y La Palma (1971). Aunque no aparecen pautas específicas que permitan adjudicar mayores probabilidades a una zona determinada del archipiélago, sí es cierto que hay una mayor actividad en las islas de Tenerife, Lanzarote y La Palma, donde se han producido las erupciones de los últimos 500 años; por contra, la isla de La Gomera no ha registrado actividad volcánica desde hace un millón de años, por lo menos.

Tampoco existe una pauta definida de la ocurrencia temporal de las erupciones; los períodos eruptivos oscilan de 1 a 237 años, con 30 años de valor promedio y una desviación típica de 25 años. No se puede prever, con los datos disponibles, cuándo podría darse un nuevo episodio eruptivo en las Canarias, aunque sí existen zonas de localización preferente. En efecto, en el vulcanismo canario, que pertenece a los basaltos alcalinos por hallarse en el interior de una placa litosférica, las erupciones son, predominantemente, de carácter fluido y, de baja explosividad, los mecanismos implicados. Los centros de emisión se alinean, por lo común, a lo largo de fracturas, abiertas ya en la corteza oceánica, donde se acumulan materiales y crecen en altura durante un intervalo de miles de años hasta configurar edificios volcánicos en tejado, que en Canarias reciben el nombre de dorsales. Se trata, en realidad, de verdaderos edificios poligénicos; a diferencia de la centralización y formación de magmas viscosos que caracterizan a la mayoría de estos volcanes, aquí son de carácter fisural y producen lavas muy fluidas. Todo el

vulcanismo de que se tenga noticia histórica se ha producido en estas dorsales, lo que facilita la labor de vigilancia.

El riesgo volcánico asociado a las dorsales, donde predominan los mecanismos eruptivos de baja explosividad, es muy bajo en comparación con el del vulcanismo explosivo que hemos visto. En realidad es casi nulo para la población, como confirma el hecho de que no se haya registrado víctimas directas de erupciones en el archipiélago. Los fenómenos típicos de las erupciones canarias son la caída de piroclastos en un radio de pocos kilómetros alrededor del centro de emisión y el flujo de coladas lávicas a favor de las pendientes topográficas. El primero no reviste peligro alguno para la población, salvo en las proximidades del volcán, y tampoco el flujo de coladas, cuyo lento avance permite una ordenada evacuación de las personas afectadas.

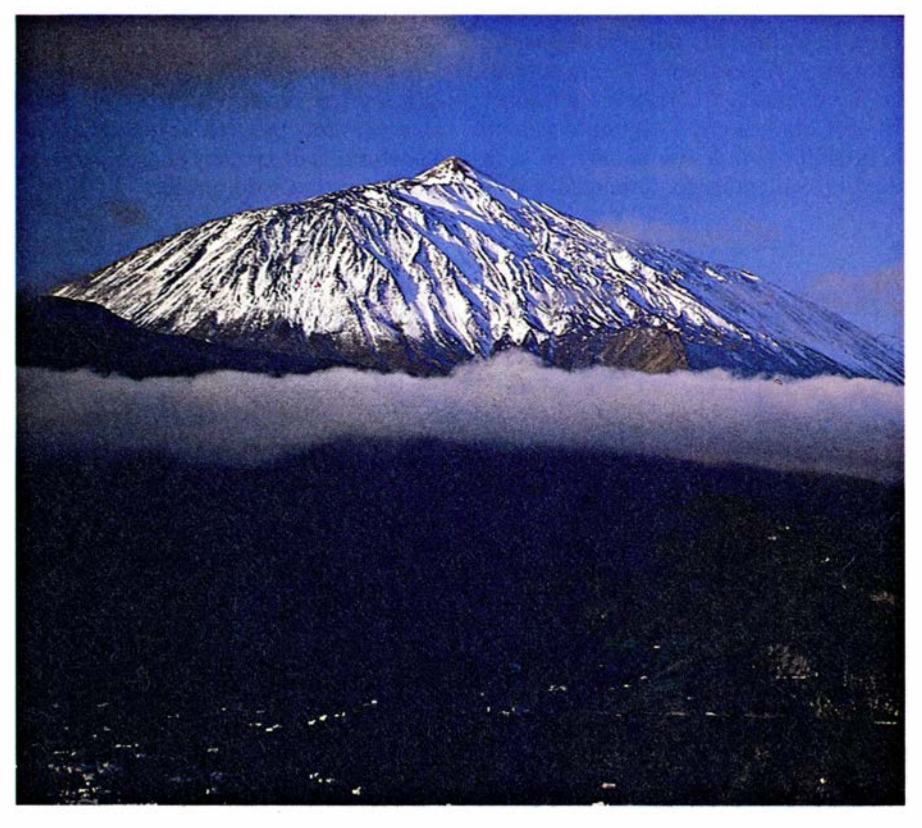
Ambos fenómenos entrañan, en cambio, un grave peligro potencial para la estructura socioeconómica de las islas, ya que pueden incendiar bosques, arrasar tierras de cultivo y viviendas, cortar las vías de comunicación, etcétera. La erupción de mayo de 1706 de la Montaña Negra, en la vertiente norte de Tenerife, no produjo víctimas, pero rellenó el puerto de Garachico, entonces el único importante y base de todo el comercio con las Indias; desde entonces, las actividades portuarias principales se trasladaron a la capital de la isla. (Se han elaborado mapas de riesgo volcánico para erupciones en las dorsales de Tenerife, que simulan, mediante ordenador, las medidas de evacuación de la población que se vería afectada en un episodio eruptivo, en función del lugar de ocurrencia.)

El panorama apuntado, de riesgo casi mínimo, se complica si tenemos en cuenta la presencia en el centro de Tenerife del complejo Teide-Pico Viejo, edificio volcánico de 3718 metros de altura, la máxima elevación del país, que se ubica exactamente en el punto de intersección de las tres dorsales de la isla (NE-SO, NO-SE y N-S). En esta zona se ha concentrado el vulcanismo en las últimas decenas de miles de años, generando un edificio de tipo central que aún permanece activo. La persistencia en un punto de la actividad volcánica ha creado una cámara magmática bastante superficial, donde el magma ha evolucionado modificando su composición hacia tipos más viscosos y ricos en substancias volátiles, con el consiguiente aumento del carácter explosivo de sus erupciones. Por otro lado, la altura alcanzada por el edificio ha incrementado su inestabilidad y permite que acumule en su cima una masa de nieve, aunque de poco espesor.

La actividad de ese complejo volcánico es muy reducida. Tuvo, sin embargo, una importante erupción en la primera mitad del siglo XV, con flujos de lava traquítica de intenso color negro por la presencia de vidrio volcánico (obsidiana), que llegaron a la costa por el norte de la isla; hay constancia de otra erupción, 3000 años atrás, que fue de carácter explosivo, aunque de muy baja intensidad, en la que se formó el edificio de la Montaña Blanca, llamada así por hallarse cubierta de pómez de ese color. Otro indicio claro de que este volcán permanece activo es la presencia de anomalías térmicas y fumarolas en su cráter —con temperaturas que alcanzan los 83 grados C, correspondiente a la ebullición del agua a esa altura—, emanación de gases volcánicos, así como frecuente actividad sísmica de origen volcánico.

Teniendo en cuenta estos factores, y ateniéndonos al concepto de riesgo volcánico —el producto de la probabilidad de que una zona se vea afectada en la ocurrencia de determinado fenómeno por las pérdidas que se esperan, económicas y en vidas humanas—, el vulcanismo basáltico fisural de las dorsales presenta una elevada probabilidad y un bajo, o incluso nulo, riesgo. Mucho mayor es el riesgo que se corre con el complejo Teide-Pico Viejo, a pesar de la escasa probabilidad de ocurrencia.

El Consejo Superior de Investigaciones Científicas de España lleva tiempo trabajando en el estudio del vulcanismo canario y la prevención de sus riesgos, aunque todavía de una manera insuficiente. En su Instituto de Recursos Naturales de la Laguna, la Estación Volcanológica de Canarias está desarrollando dos programas, fruto de la colaboración entre ese organismo y el gobierno del archipiélago, gracias a los cuales se ha instalado la primera estación sísmica con enlace telemétrico en la base del Teide y se analizarán periódicamente parámetros vulcanológicos válidos como precursores: contenido de radón y otros gases volcánicos en las aguas de galerías subterráneas, variaciones de temperaturas en zonas de anomalía térmica y otras magnitudes. Las técnicas reseñadas permitirán un mayor conocimiento del vulcanismo activo de esa área densamente poblada y una mejor preparación frente al riesgo de erupciones.



9. EL TEIDE, en la isla canaria de Tenerife, es el único estratovolcán activo y la máxima elevación de España, con sus 3718 metros sobre el nivel del mar. Aunque su historia volcánica reciente nos lo muestre poco activo (la última erupción se produjo en la primera mitad del siglo XV), su altitud, mayor inestabilidad, presencia de nieve buena parte del año y posibilidad de mecanismos eruptivos violentos, aconsejan vivamente su estudio y vigilancia. El gobierno de Canarias y el CSIC —a través de su Estación Volcanológica de Canarias— han desarrollado un programa que ha permitido la instalación de una primera estación sísmica con enlace telemétrico en la base del volcán, que aporta información en tiempo real de su nivel de actividad.

Ciencia y sociedad

Corriente pulsante

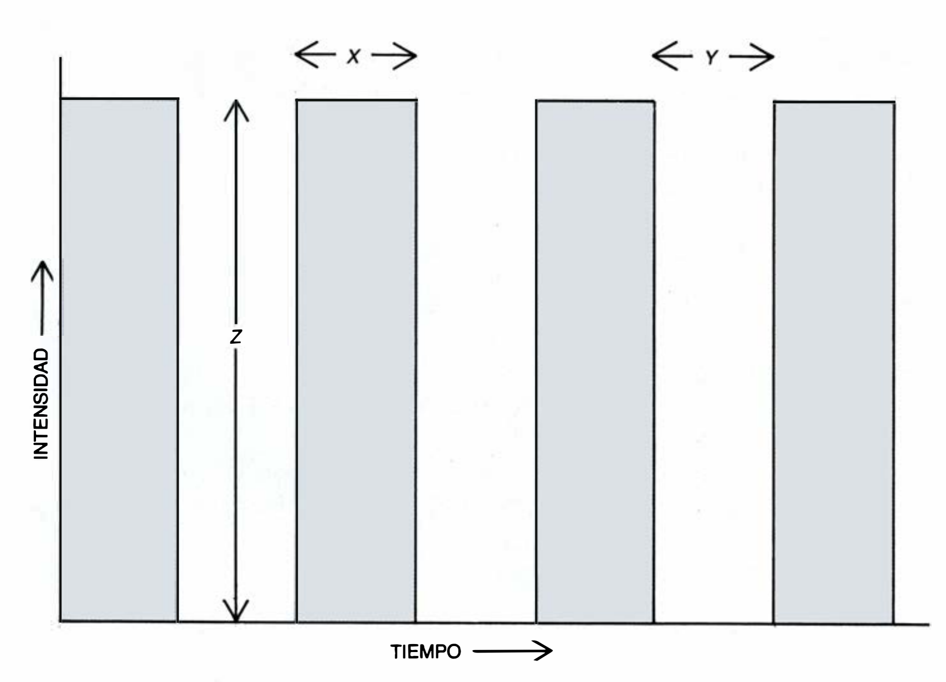
En la técnica habitual de electrodeposición de películas metálicas sobre una base plástica o metálica se utiliza corriente eléctrica continua, obtenida, mediante los oportunos rectificadores, de la corriente alterna que sirve la red. Quedó descartado en principio el empleo directo de la propia corriente alterna, porque el metal que se electrodepositase durante la fase catódica se disolvería en la anódica, sin que llegara a formarse recubrimiento alguno.

En estudios de electrodeposición con corriente alterna realizados en la década de 1940, en los que se modificaba la duración de la fase anódica con respecto a la fase catódica, se obtuvieron depósitos metálicos con un alto grado de brillo y reducida porosidad. Parecía iniciarse de ese modo una prometedora aplicación de la corriente alterna en la electrodeposición de metales.

En la década de 1950, de modo independiente y sirviéndose de dispositivos diferentes, los rusos Krivtsov y Ozerov utilizaron, para la deposición de níquel, una variante de esa técnica, la denominada "corriente pulsante". El elevado costo del equipo desalentó la introducción de esa técnica en la industria. Sin embargo, a mediados de la década de los setenta, los avances tecnológicos abarataron su aplicación y resurgieron con notable ímpetu los estudios sobre el nuevo procedimiento, que, tímidamente, empezó a abrirse paso en los talleres galvanotécnicos.

En la electrodeposición de metales por corriente pulsante, o por impulsos, una corriente eléctrica de onda cuadrada atraviesa el electrólito que contiene el metal que debe formar el recubrimiento. Las corrientes de onda cuadrada se caracterizan porque el tiempo de ascenso y caída de sus impulsos está comprendido entre 0,01 y 7,5 milisegundos, a frecuencias básicas de entre 200 y 5000 hertz. Los generadores de ese tipo de corriente "abren" el paso a la electricidad durante un período de tiempo predeterminado y seguidamente lo "cierran" durante otro lapso, también estipulado, repitiéndose a continuación el ciclo. La relación entre el tiempo de apertura (*X* en la figura 1) y el de cierre (*Y*) define el ciclo de trabajo, o ciclo de función (*Z*), que equivale al cociente entre el tiempo de apertura y la suma de los tiempos de apertura y de cierre.

Los rectificadores de pulso modernos suelen llevar acoplado un panel que permite programar el tiempo de apertura, el de cierre y el pico, o cresta, de intensidad o de tensión.



1. Parámetros de electrodeposición con corriente pulsante

Los efectos principales de la corriente pulsante son la reducción del potencial catódico y el adelgazamiento de la fina película constituyente de la capa de difusión, lograda sin necesidad de empobrecer dicha capa en iones del metal. En efecto, cuando pasa la corriente (durante el tiempo de apertura), se agotan rápidamente los iones del metal próximos al cátodo, que se depositan sobre él, registrándose un enriquecimiento de la película en iones no metálicos.

En el instante siguiente, al interrumpirse la corriente (durante el tiempo de cierre), los iones metálicos de la zona que rodea al cátodo se difunden prestamente al interior de la película catódica, manteniéndose de ese modo la concentración de iones metálicos de la capa de difusión, mientras las burbujas de gas hidrógeno y las impurezas adsorbidas en el cátodo retornan a la masa del electrólito.

El ciclo se repite una y otra vez a lo largo del proceso de electrólisis, obteniéndose recubrimientos metálicos de calidad superior a los que rinde la técnica convencional con corriente continua, en especial cuando se trata de la deposición de metales preciosos sobre piezas de las que se esperan altas prestaciones técnicas.

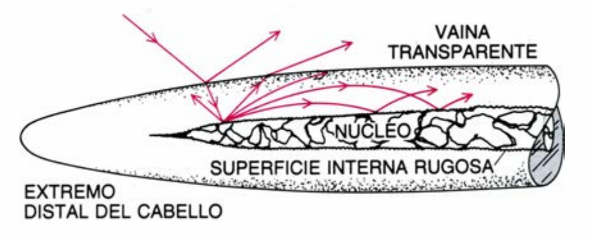
Desde 1974, cuando empezó su aplicación a la industria, hasta hoy, la electrodeposición mediante corriente pulsante ha cobrado gran auge y ha sido tema de numerosas investigaciones empeñadas en mejorar su conocimiento y utilización. Así se puso de manifiesto en el Primer Simposio Internacional sobre Electrodeposición Pulsante, celebrado en Boston en 1979, y en el tercero, que acogió Washington en octubre de 1986 y donde se presentaron numerosos trabajos sobre la aplicación de esa técnica a los campos de la telecomunicación, los semiconductores, electrónica e industria aeroespacial. En este último, el procedimiento se ha demostrado de gran utilidad en la aplicación de películas o aleaciones de oro sobre material de titanio, de acero inoxidable y de Inconel, utilizados en componentes de satélites artificiales. El bajo valor de emisión térmica, la alta reflectividad a los rayos infrarrojos, la escasa porosidad y el elevado brillo de los recubrimientos de oro obtenidos

por esa técnica han permitido su aplicación a ese fin, que muy probablemente se extenderá en un futuro próximo a otros componentes de la industria aeroespacial. (Enrique Julve.)

Osos solares

En días despejados, a temperaturas ambientales de cero grados Celsius, un captador de energía solar de muy buenas prestaciones convierte la radiación en calor con una eficacia máxima del 40 por ciento. Existe un captador natural que convierte en calor parte del espectro de la radiación solar con eficacias superiores al 95 por ciento: el pelaje del oso polar.

Esa capacidad del oso polar se descubrió hace unos siete años, de modo indirecto, cuando el gobierno canadiense puso en marcha un estudio de poblaciones de la foca ártica. Ante la dificultad de distinguirlos del fondo blanco del invierno ártico, se supuso que el calor corporal que despiden revelaría, en películas infrarrojas, la presencia de los animales. Pero, el excelente aislamiento natural de las focas,





2. Pelo de oso polar, que podría constituir un cable de fibra óptica natural. El núcleo, rugoso, está rodeado por una vaina maciza que, según parece, atrapa la luz ultravioleta y la dirige hacia la piel

que impide la pérdida de calor, echó por tierra los planes.

El problema de la localización lo resolvió la fotografía ultravioleta. Mientras que el entorno reflejaba la radiación correspondiente a la región ultravioleta del espectro, el pelaje de las focas la absorbía. En investigaciones ulteriores se comprobó que también absorbían esa radiación otros pelajes de mamíferos árticos, y que el del oso polar era el que lo hacía con mayor eficacia.

Ante esos datos, Richard E. Grojean, de la Universidad del Noroeste y de la Comandancia Natick de Investigación y Desarrollo del Ejército de los Estados Unidos, pensó que el estudio del pelaje del oso polar podría ayudar al hombre en su empeño por domeñar la energía de la radiación solar. Grojean descubrió que los pelos del oso no eran blancos, sino incoloros, y que, vistos al microscopio, semejan fibras de cuarzo. El núcleo central del pelo dispersa la radiación que le llega, confiriéndole la apariencia blanca. (Un efecto dispersor similar explica la blancura de los copos de nieve.) Grojean sostiene que el cañón del pelo conduce la radiación dispersada hasta la superficie de la piel (que en realidad es negra),




3. *Vaso calcolítico* (izquierda) *procedente del yacimiento de "El Dorado" (Córdoba), cuyo interior contenía las cenizas de un individuo de edad juvenil; de él se recuperaron las coronas de la dentición mandibular. Y dentición mandibular* (derecha) *extraída del vaso calcolítico. Carece de los segundos incisivos y de los terceros molares; perteneció a un joven que murió entre los 13 y los 15 años*

donde se absorbe y se convierte en calor. Ese calor generado pasivamente complementa el de origen metabólico.

Las aplicaciones del fenómeno parecen prometedoras. Según Gregory Kowalski, colega de Grojean en la Noroeste, él ha comprobado un incremento del 50 por ciento de la eficacia de captadores solares planos al recubrirlos de fibras piliformes. Si se modificaran los paneles solares hasta captar una porción mayor del espectro solar, su rendimiento mejoraría de golpe, afirma Kowalski. Gran parte de la radiación, en especial la ultravioleta, atraviesa la cobertura nubosa y llegaría hasta los captadores solares incluso en días nublados. El sistema ahorraría también la necesidad de utilizar complejos dispositivos de seguimiento para mantener las placas encaradas al sol, pues los aparatos recubiertos de "pelos" capturan la radiación que procede de todas las direcciones.

Incineración en el Calcolítico español

Ha crecido en los últimos años el interés por desentrañar las pautas etológicas funerarias de nuestros antepasados del Calcolítico. Los enterramientos practicados en Andalucía occidental durante la edad del Cobre son de seis tipos, cinco artificiales —cista megalítica, galería cubierta, sepulcro de corredor, tholos y cueva artificial—, y uno practicado en el interior de grutas naturales. Consiste la cista megalítica en una cámara simple, de forma rectangular, construida con grandes losas de piedra y cobertura adintelada. La galería cubierta incluye sepulcros en los que la cámara y el corredor no están diferenciados o, si lo están, la separación es pequeña; se construyen con grandes losas de piedra y cubierta adintelada. Los sepulcros con cámara y corredor perfectamente diferenciados, siempre que la cámara no sea circular, se denominan sepulcros de corredor; se hicieron con losas grandes de piedra y cobertura adintelada. El tholos es un sepulcro de cámara circular o tendiente a ella, cubierto de falsa cúpula, con corredor o sin él; la técnica constructiva es de losas grandes o mampostería. En la cueva artificial, una tumba excavada en el suelo, se emplearon, a veces, grandes losas.

En los sepulcros estudiados hasta el momento, se ha comprobado una mayor abundancia de galerías cubiertas y de cuevas artificiales sobre las restantes formas, seguidas por el tholos, aunque en un porcentaje notablemente inferior. Es el tholos un sepulcro propio de terrenos llanos, con buenas tierras de labor; los construidos en estas zonas alcanzan mayor tamaño y perfección que los de zonas montañosas.

El tholos aparece en Betaña en el quinto milenio a. de C. Cuando se desarrolla en la península Ibérica presenta ya una rica variedad de elementos exóticos: vasos de piedra y hueso, ídolos, sandalias de marfil procedentes, probablemente, del norte de Africa, ámbar, quizás obtenido como intercambio comercial, piezas en cobre y oro, etcétera. Todo esto permite suponer que las sociedades de esta época poseían un potencial económico que les permitía obtener objetos de lujo.

Durante el período calcolítico, los enterramientos humanos consisten en simples inhumaciones donde el difunto es sepultado con evidentes muestras de un ritual. Esta práctica está ocasionalmente acompañada con la presencia de fuego sobre los cuerpos muertos, sin que existan claros indicios de la práctica crematoria o incineración del difunto durante la edad del Cobre, costumbre que será habitual en culturas ulteriores. Aunque conociendo este hecho, uno de nosotros (Cabrero) ha estudiado exhaustivamente el yacimiento calcolítico de "El Dorado" (Córdoba), donde apareció —entre muchos indicios que confirman la tesis sostenida en este artículo— un vaso calcolítico en



4. *Cráneos de la Edad del Cobre descubiertos en el yacimiento de "Alcalá del Valle" (Cádiz). El de la izquierda corresponde a un adulto, varón; el de la derecha mantiene abiertas las suturas craneales, de donde se infiere que murió siendo adulto joven, y pertenece a una mujer*

cuyo interior existían cenizas humanas que contenían las coronas de los dientes de un individuo de edad juvenil. Conscientes de la rareza de este evento, que sería único en el suroeste de España, y teniendo en cuenta que el vaso calcolítico fue extraído de un monumento excavado en los años cincuenta, no excluimos la posibilidad de que el vaso fuera reutilizado por culturas posteriores donde la cremación era ya una práctica habitual. No obstante, los estudios arqueológicos del momento nos inducen a pensar que es una cremación calcolítica, pues no existen tampoco indicios de culturas más recientes a la Edad del Cobre en todo el yacimiento, que consta de un poblado y de tres tumbas.

En el suroeste de España existen referencias bibliográficas que recogen la acción del fuego sobre los difuntos. Consta que en el interior de algunos sepulcros encendían hogueras sin que afectaran a los esqueletos ni a los ajuares, como sucedió con el "Dolmen de Soto", "Zarcita" y "Cueva del Vaquero". También consta la existencia de un cráneo trepanado en vida con uso del fuego, en auxilio de una intervención quirúrgica. En otras necrópolis andaluzas se han descubierto restos de esqueletos quemados, como en "Los Millares", "Los Liniales 9", "Jautón 5" y otras sepulturas almerienses; presentan quemados algunos huesos. En la necrópolis del "Barranquete" (Almería) se encontraron enterrados en la cámara unos cincuenta cadáveres, algunas de ellos calcinados. En "La Encantada II" había huesos esparcidos por la cámara, algunos de ellos quemados, mientras que los cadáveres contenidos en los nichos, unos ocho, estaban inhumados. "La Encantada III" contenía, en un hoyo de la cámara, restos de unos sesenta cadáveres.

Numerosos sepulcros megalíticos de la región de Gorafe (Granada) albergaban huesos carbonizados. Recientemente se han descubierto abundantes huesos quemados en muy diverso grado en una tumba de la necrópolis de "Murviedro" (Murcia), como si hubieran sufrido una incineración parcial que afectaba en mayor medida al cráneo, circunstancia coincidente con el "Barranquete". De la referida tumba de "Murviedro" proceden huesos probablemente quemados en el interior de algunas vasijas.

El recipiente cerámido de "El Dorado" apareció en la cámara del sepul-

cro megalítico. Posee forma globular y contenía las cenizas de un individuo quemado, entre las que se recuperó, casi completa, la dentición mandibular, reducida, por la acción del fuego, exclusivamente al esmalte de las coronas dentales. El estudio antropológico de la dentición ha sido realizado por uno de los autores (Alcázar). La mandíbula fue introducida en el interior del vaso sobre una capa de cenizas, con el mentón dirigido hacia el interior. Sobre ella se vertió otra capa adicional de cenizas hasta cubrirla totalmente. El estado del desgaste dental y las piezas aparecidas sugieren que la dentición corresponde a un individuo joven, que murió entre los trece y los quince años de edad. El vaso crematístico de "El Dorado" ha sido datado por Cabrero entre el 2500 y el 1800 a. de C. Representa el primer hito de esta práctica ritualista en el suroeste español. Este y los casos anteriormente descritos reflejan las costumbres y creencias de una sociedad calcolítica que construía sepulcros megalíticos para enterrar a sus difuntos.

Aunque escasean los estudios realizados sobre la edad de fallecimiento de las poblaciones calcolíticas en el sur peninsular, los resultados a que llegan son coincidentes. Así, de los nueve individuos enterrados en el sepulcro del "Moral" (Málaga) la mayoría poseía una edad entre 20 y 40 años, siendo excepcional uno de 10 años y dos entre 45 y 60 años. De la cueva artificial de "Cabra" (Córdoba) proceden siete personas, cuatro de edad entre 50 y 65 años, dos entre 20 y 30 años y otro juvenil. En el "Barranquete" (Almería), de 107 esqueletos se observó que la mortalidad ascendía progresivamente desde las primeras edades hasta hacerse muy grande en la edad adulta; de hecho, el 59,8 por ciento pertenecen a esta etapa. En el yacimiento de "Alcalá del Valle" (Cádiz), de 16 individuos, 10 pertenecen al calcolítico; entre ellos, 6 son adultos jóvenes, 2 son adultos y otros 2 son seniles.

Deducimos como conclusión que la vida media de los pobladores del sur de España era, durante la edad del Cobre, bastante menor que la actual. La longevidad era rara, así como, probablemente, la mortalidad en la primera etapa de la infancia. (José Alcázar y Rosario Cabrero.)

Nueva luz

La supernova 1987A se ha apagado considerablemente, cuando menos en el espectro visible, desde su fulgurante aparición en la Gran Nube de Magallanes, el 24 de febrero del año pasado. Pero su contribución a la astronomía resulta cada vez más brillante.

Las observaciones probablemente más satisfactorias constituyen la primera confirmación de la teoría según la cual los elementos pesados —del silicio en adelante— que forman gran parte de la tierra y de sus habitantes vivos se forjan en las estrellas en explosión. Los primeros indicios de que en el interior de 1987A se está produciendo la nucleosíntesis de elementos pesados se han obtenido de su correspondiente "curva de luz". Durante las primeras semanas que siguieron al fogonazo inicial, la supernova fue palideciendo; muchos teóricos creían que proseguiría en esa pauta.

Otros, en particular Stanford E. Woodsley, de la Universidad de California en Santa Cruz, predijeron un comportamiento distinto. Woosley había construido un modelo de supernova en el cual la explosión inicial genera grandes cantidades de cobalto 56; ese isótopo, que tiene una vida media de 77 días, se desintegra emitiendo rayos gamma y forma hierro estable. Woosley propuso que la supernova brillaría de nuevo cuando su capa externa se expandiera hasta una distancia suficiente para que escaparan los rayos gamma, atenuados a luz visible. En efecto, la supernova brilló desde mediados de marzo hasta finales de mayo; desde entonces ha palidecido a un ritmo que concurda con el de la desintegración del cobalto 56.

Woosley y otros teóricos predijeron también que, cuando la cubierta externa se hiciera más transparente aún, se detectaría mejor la desintegración del cobalto, primero en forma de rayos X y, por último, de rayos gamma. En octubre, los instrumentos situados a bordo de la nave espacial soviética *Mir* y del satélite japonés *Ginga* captaron los rayos X. Las semanas siguientes, cuatro detectores financiados por la NASA (uno a bordo del satélite *Misión Máximo Solar* y los otros en globos lanzados a la atmósfera superior sobre Australia) recibieron los rayos gamma.

El Observatorio Aéreo Kuiper, de la NASA, un avión equipado con sensores infrarrojos, ha obtenido más pruebas de la riqueza en elementos de la supernova. Según Fred C. Gillett, de la NASA, en dos vuelos realizados en noviembre, el observatorio detectó emisiones infrarrojas típicas del cobalto, níquel, silicio, azufre y argón.

Todos esos resultados, señala Edward L. Chupp, de la Universidad de New Hampshire, jefe de investigación del proyecto *Máximo Solar*, representan "un triunfo real de la teoría". Opina que, en particular, la detección de los rayos gamma es de importancia comparable a la de neutrinos cuando apareció la supernova, que confirmó las teorías sobre el comienzo de esas explosiones.

Los científicos confían en que las observaciones futuras les ayuden a revelar la historia de la estrella precursora de 1987A: Sanduleak −69 202, una gigante azul. Se creía que esas estrellas eran demasiado calientes y jóvenes para explotar. Sin embargo, algunos investigadores mantienen que Sanduleak fue antes una gigante roja que, hace miles de años, se desprendió de sus capas exteriores en forma de viento solar. Robert P. Kirshner, del Centro de Astrofísica Harvard-Smithsoniano, considera que los datos obtenidos por el satélite *Explorador Ultravioleta Internacional* apoyan su interpretación: muestran que el fogonazo ultravioleta de la supernova caldea un gas, rico en nitrógeno y que se mueve más lentamente, supuesto componente de las capas exteriores de las gigantes rojas.

Los científicos pretenden determinar asimismo si la explosión de Sanduleak dejó algún resto. Los neutrinos que acompañaban a la explosión sugieren que el núcleo se hundió y formó una estrella de neutrones. Si la estrella de neutrones gira con velocidad suficiente, podrá detectarse constituída en púlsar, un emisor de fogonazos de energía electromagnética que se suceden con gran periodicidad. Thomas A. Prince, del Instituto de Tecnología de California, señala, sin embargo, que el que hasta la fecha no se haya detectado el púlsar establece un límite superior a su intensidad.

Las observaciones más intrigantes quizá sean las de un equipo de Harvard que aplica la técnica de interferometría de mácula. En junio de 1987, el grupo anunció que había avistado una misteriosa "compañera" de la supernova (véase "Ciencia y sociedad", agosto de 1987). En observaciones posteriores, esos científicos no han detectado la compañera. Recientemente han informado de que, medido el diámetro de la supernova, lo han encontrado cinco veces mayor que el valor teóricamente previsto. Robert V. Stachnik, de la NASA, que conoce de cerca el trabajo del grupo, insiste en que debe tomarse en serio la observación.

Evolución del clima en los planetas terrestres

Se creyó antaño que los planetas con climas templados, semejantes al de la Tierra, eran raros en nuestra galaxia. Los modelos matemáticos sugieren que, si hay planetas fuera del sistema solar, muchos podrían ser habitables

James F. Kasting, Owen B. Toon y James B. Pollack

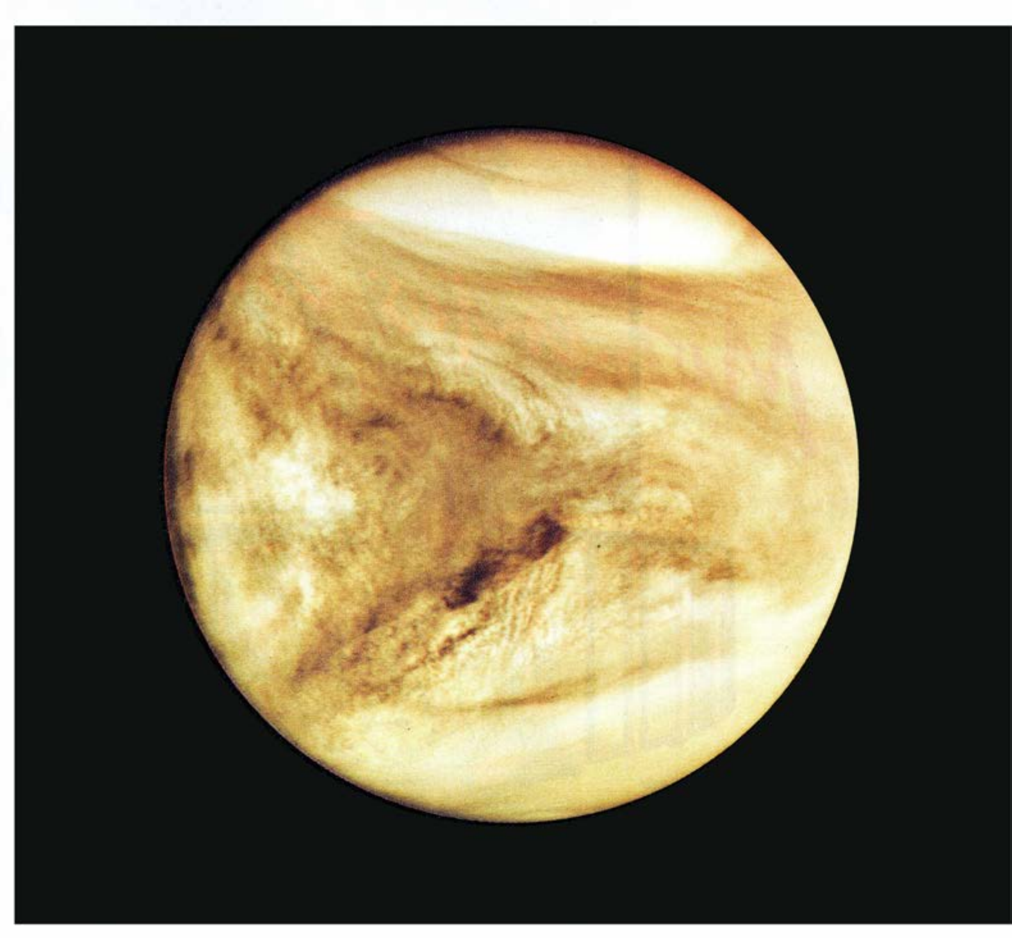
¿Por qué es Marte demasiado frío para albergar la vida, Venus demasiado cálido y la Tierra ideal? A primera vista, la respuesta al problema de Goldilocks, como se conoce esa cuestión en climatología, parece sencilla. El sentido común sugiere que la Tierra, con una temperatura media de 15 grados Celsius, se formó casualmente a la distancia correcta del Sol, mientras que ni Marte (−60 grados Celsius) ni Venus (460 grados Celsius) tuvieron la misma suerte. Sólo en la superficie de la Tierra se encuentra, pues, el agua en forma líquida, que es crucial para la vida.

Pero no debemos atribuir a la casualidad toda la explicación de las temperaturas de estos planetas terrestres, o rocosos. En nuestra opinión, los tres vecinos, formados tras la colisión mutua de grandes cantidades de cuerpos conocidos como planetésimos, fueron antaño semejantes en muchos aspectos. Presentaban, en su superficie, minerales parecidos y, en su atmósfera, gases similares (entre ellos el dióxido de carbono y el vapor de agua); los tres, además, conocían un clima templado que les permitía mantener agua líquida en su superficie. Si adquirieron luego climas espectacularmente distintos, debióse, en gran parte, a su diferente capacidad para crear un ciclo de dióxido de carbono entre la corteza y la atmósfera. El dióxido de carbono, como el vapor de agua y otras substancias, es un gas "invernadero": permite el paso de la radiación solar a través de él, pero absorbe la radiación infrarroja (calorífica) que procede del planeta y reemite de nuevo parte de este calor hacia su superficie.

En concreto, los cálculos realizados por el grupo de los autores en la Administración Nacional de Aeronáutica y del Espacio (NASA) y por otros investigadores sugieren que la Tierra ha gozado siempre de un clima moderado en virtud de la siguiente razón principal: su mecanismo cíclico aumenta la cantidad de dióxido de carbono en la atmósfera cuando la superficie del planeta se enfría y reduce dicha cantidad cuando aumenta la temperatura en la superficie. Marte está helado porque ha perdido la capacidad de desprender de nuevo el gas a su atmósfera; Venus es un infierno porque experimentó el problema opuesto: no tiene manera de extraer dióxido de carbono de su atmósfera. Mercurio, el otro planeta terrestre, carece de atmósfera; su temperatura está controlada exclusivamente por el Sol.

La paradoja de un Sol débil y joven

Nuestro interés en el papel desempeñado por el dióxido de carbono en la evolución de la Tierra, Marte y Venus tiene su origen en otro rompecabezas cosmológico relacionado con el origen



1. VENUS, LA TIERRA Y MARTE, (*de izquierda a derecha*), representados aproximadamente a escala. Los tres pudieron en su historia inicial gozar de un clima templado que les permitiera conservar la vital agua

de la Tierra: la paradoja de un Sol joven y débil. En su inmensa mayoría, los modelos de evolución estelar indican que el Sol era entre un 25 y un 30 por ciento más débil cuando se formó el sistema solar, hace 4600 millones de años, de lo que es hoy. Desde entonces, la luminosidad, o intensidad solar, ha crecido, de forma bastante lineal según parece, con el tiempo.

Como señalaron quince años atrás Carl Sagan y George H. Mullen, de la Universidad de Cornell, la paradoja surge del planteamiento siguiente: si la primitiva atmósfera terrestre tenía la misma composición que ahora, un Sol débil habría promovido una Tierra recubierta de hielos hasta hace unos 2000 millones de años; pero nuestro planeta no se congeló. La verdad es que, de acuerdo con la información ofrecida por las rocas sedimentarias, la Tierra ha tenido océanos líquidos desde hace al menos 3800 millones de años, cuando comienza el registro geológico. Además, la vida ha estado presente durante 3500 millones de años, por lo menos, lo que demuestra que la superficie terrestre nunca ha estado enteramente helada durante ese tiempo. (El agua puede permanecer fluida mientras la temperatura esté comprendida entre cero y 374 grados Celsius; hierve y se evapora a 100 grados Celsius al nivel del mar, pero permanecería líquida a temperaturas más altas si aumentara la presión atmosférica.)

Sagan y Mullen comprendieron que la paradoja desaparecía si se suponía que la atmósfera terrestre había ido cambiando en el curso del tiempo. Por ejemplo, si el joven planeta hubiera tenido menos nubes de las que hay hoy, habría reflejado de nuevo al espacio menos radiación solar incidente sobre el planeta, y éste se hubiera mantenido, en consecuencia, más caliente. Alrededor de un 30 por ciento de la radiación solar que ahora llega a la cima de la atmósfera se devuelve al espacio, la mayor parte de ella por reflexión en las nubes. Una Tierra más fría habría tenido menos nubes, pero el registro geológico aboga por una Tierra primitiva más caliente que la que conocemos. Parte del planeta está hoy cubierta de glaciares; no hay ninguna señal de glaciaciones similares precedentes a 2700 millones de años atrás.

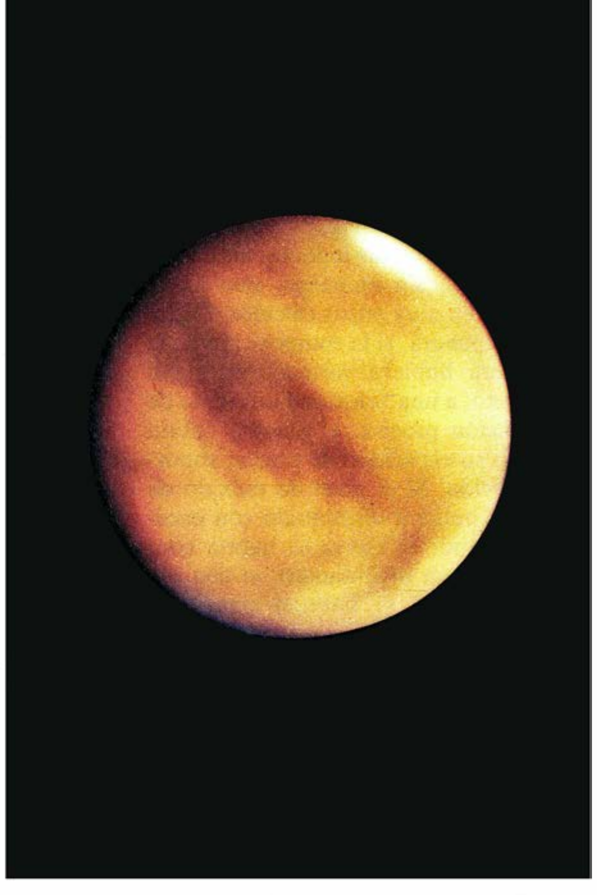
La explicación más plausible es que el efecto de invernadero fuera más intenso en un pasado remoto. Sagan y Mullen sugirieron que el amoniaco (NH_3), absorbente eficaz del infrarrojo, pudo haber conducido a un clima más cálido si el gas hubiera constituido apenas 100 de cada millón de moléculas de aire. Pero estudios posteriores han demostrado que el Sol hubiera convertido rápidamente el amoniaco en nitrógeno e hidrógeno, que no contribuyen al invernadero, a menos que la superficie del planeta lo estuviera suministrando sin cesar a la atmósfera.

El dióxido de carbono

Otras investigaciones se han centrado en el dióxido de carbono, que la radiación solar no descompone fácilmente. El dióxido de carbono abunda aquí; la cantidad ahora almacenada en el planeta, en forma de rocas carbonatadas, ejercería una presión de unos 60 bares si se desprendiera a la atmósfera. (Un bar equivale a la presión atmosférica al nivel del mar. La presión ejercida por el dióxido de carbono contenido en la atmósfera actual se cifra en unos 0,0003 bar.) Si sólo hubiera habido unos pocos decibares del dióxido de carbono almacenado en forma gaseosa, su efecto de invernadero hubiera producido un calentamiento capaz de compensar la disminución de radiación solar.

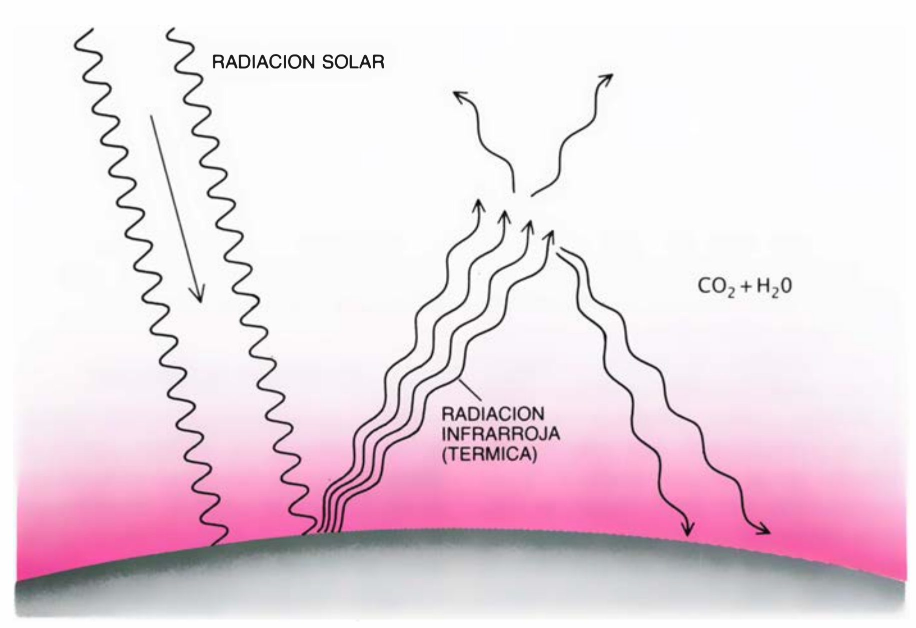
La idea de que concentraciones ma-







en fase líquida en su superficie. Los modelos de ordenador sugieren que la diferente capacidad para realizar el ciclo del dióxido de carbono entre la atmósfera y el suelo —y no solamente la distancia al Sol— facilitó que Venus perdiera su agua y Marte se congelara, mientras la Tierra permanecía habitable.



2. EFECTO INVERNADERO, que se produce cuando ciertos gases, sobre todo el dióxido de carbono y el vapor de agua, calientan la superficie de un planeta considerado. Tales gases permiten el paso de la radiación solar hasta la superficie del planeta, pero interceptan la radiación infrarroja (térmica) que el planeta emite hacia el espacio, y reenvían gran parte de esta energía hacia la superficie. Estos gases elevan la temperatura en la superficie terrestre unos 35 grados Celsius por encima de la temperatura que habría si faltaran.

yores de dióxido de carbono pudieron haber protegido a la Tierra primitiva de su congelación dio pronto lugar a otra noción afín: si el nivel de dióxido de carbono disminuyese a una velocidad que compensase el aumento de la radiación solar con el tiempo, tal disminución podría explicar que la temperatura de la Tierra persistiera siempre dentro de límites razonables. Un investigador de la NASA, Michael H. Hart, emprendió el cálculo de tal velocidad compensadora.

Hart obtuvo una solución en la que los niveles del gas disminuían de forma aproximadamente logarítmica con el tiempo. Su descubrimiento más interesante fue, sin embargo, que había acertado en muy pocos cálculos. Dicho de otra manera, si la composición de la atmósfera hubiera variado, en algún momento, a una velocidad diferente de su solución precisa, el planeta habría perdido su capacidad de prestar soporte a la vida. Si el nivel de dióxido de carbono hubiera disminuido con excesiva lentitud, la Tierra se habría convertido en un invernadero; si con excesiva rapidez, se habrían congelado los océanos.

Hart realizó cálculos similares para casos en que la distancia entre la Tierra y el Sol variase ligeramente. Halló que si la Tierra se hubiera formado a una distancia del Sol un 5 por ciento menor, la atmósfera se habría calentado tanto que los océanos se habrían evaporado; se hubiera producido el efecto de invernadero descontrolado. Por el contrario, el planeta habría sufrido una glaciación sin control si se hubiera formado tan sólo un 1 por ciento más lejos del Sol. Unicamente en el estrecho dominio de órbitas que se desenvolviesen entre 0,95 y 1,01 unidades astronómicas podía evitarse una u otra de esas catástrofes climáticas. (Una unidad astronómica es la distancia media entre la Tierra y el Sol, es decir, 149,6 millones de kilómetros.) Hart denominó a esta banda de distancias orbitales "zona de habitabilidad continua" (ZHC).

Las conclusiones de Hart nos ponían en un aprieto: daban a entender que la Tierra hubo de superar extraordinarias dificultades para evitar el sino de Marte o de Venus. Los investigadores no pudieron descubrir el fallo de esta hipótesis hasta hace poco. De acuerdo con un modelo matemático elaborado por James C. G. Walker y Paul B. Hays, de la Universidad de Michigan, y por uno de los autores (Kasting), las variaciones de la concentración de dióxido de carbono no se produjeron por mero azar, sino que fluctuaron en respuesta a determinadas variaciones experimentadas por la temperatura de superficie. Cuando la temperatura subía, los niveles de dióxido de carbono atmosféricos decrecieron, enfriando la superficie; cuando ésta se enfriaba, la concentración de dióxido de carbono atmosférico aumentaba y calentaba la superficie. La existencia de este bucle de realimentación negativa implica que la Tierra probablemente nunca corrió peligro de sufrir ni el efecto invernadero descontrolado, ni la glaciación ilimitada postulados por Hart.

En el sistema de realimentación propuesto, el papel intermedio corresponde al ciclo geoquímico del carbonato-silicato, que da cuenta de alrededor de un 80 por ciento del dióxido de carbono que se ha intercambiado entre la parte sólida de la Tierra y su atmósfera a lo largo de un periodo de tiempo superior a los 500.000 años. El ciclo comienza cuando el dióxido de carbono de la atmósfera se disuelve en el agua y forma ácido carbónico (H_2CO_3). La lluvia erosiona rocas que contienen minerales de calcio y silicatos (compuestos de calcio, silicio y oxígeno). En tal proceso, el ácido carbónico reacciona químicamente con las rocas, liberando iones de calcio y bicarbonato (Ca^{++} y HCO_3^-) en el agua contenida en el suelo. Esta transporta los iones hacia los arroyos, ríos y, en último término, hasta los océanos.

El ciclo del carbonato-silicato

En el mar, el plancton y otros organismos incorporan los iones en las conchas de carbonato cálcico ($CaCO_3$). Cuando los organismos mueren, caen al fondo del océano, donde forman sedimentos de carbonato. Con el transcurso de los milenios, el fondo de los mares se expande, transportando esos sedimentos a los márgenes de los continentes. Allí, el fondo de los mares se desliza bajo las masas terrestres y se pliega, hundiéndose hacia el interior del planeta.

A medida que el sedimento queda sometido a este hundimiento o subducción, y a temperaturas y presiones más altas, el carbonato cálcico reacciona con el silicio (cuarzo), volviendo a formar rocas de silicatos (proceso conocido como metamorfismo de los carbonatos) y desprendiendo dióxido de carbono gaseoso. El gas vuelve entonces a entrar en la atmósfera por medio de fisuras del fondo de los océanos o, más violentamente, mediante erupciones volcánicas en la vecindad de los bordes de las placas tectónicas.

Walker y sus colegas advirtieron que las variaciones de la temperatura superficial en el curso del tiempo deberían afectar a la cantidad de dióxido de carbono presente en la atmósfera, y ésta, a su vez, al calentamiento provocado por el efecto de invernadero. Supongamos que la temperatura en superficie descendiera, por ejemplo a causa de una disminución de la energía emitida por el Sol. Cuando la temperatura del océano desciende, se evapora a la atmósfera menos vapor de agua, llueve menos y se produce, por tanto, menos erosión. Bajo tales cir-

cunstancias, disminuye la velocidad a la que el dióxido de carbono abandona la atmósfera, mas no la velocidad a la cual el dióxido de carbono se regenera por el metamorfismo de los carbonatos y se desprende a la atmósfera. El resultado neto es una acumulación del gas en la atmósfera, un aumento en el calentamiento de invernadero y, en consecuencia, un retorno a temperaturas de superficie más altas.

Por el contrario, si aumentase la temperatura en la superficie, la velocidad de evaporación de los océanos aumentaría y, con ella, la cantidad de lluvia. Se intensificaría la erosión de las rocas silíceas y, por tanto, la eliminación del dióxido de carbono de la atmósfera. Disminuiría entonces el calentamiento de invernadero.

Ilustraremos mejor la realimentación atendiendo a un caso extremo. Si los océanos se helaran por completo, cesaría la producción de lluvia y se acumularía el dióxido de carbono en la atmósfera. A la actual velocidad de desprendimiento del gas, en 20 millones de años, tiempo insignificante desde el punto de vista geológico, se alcanzaría en la atmósfera la presión de un bar de dióxido de carbono. Esta cifra elevaría la temperatura de superficie en unos 50 grados, más que suficiente para fundir los hielos y restablecer condiciones climáticas moderadas.

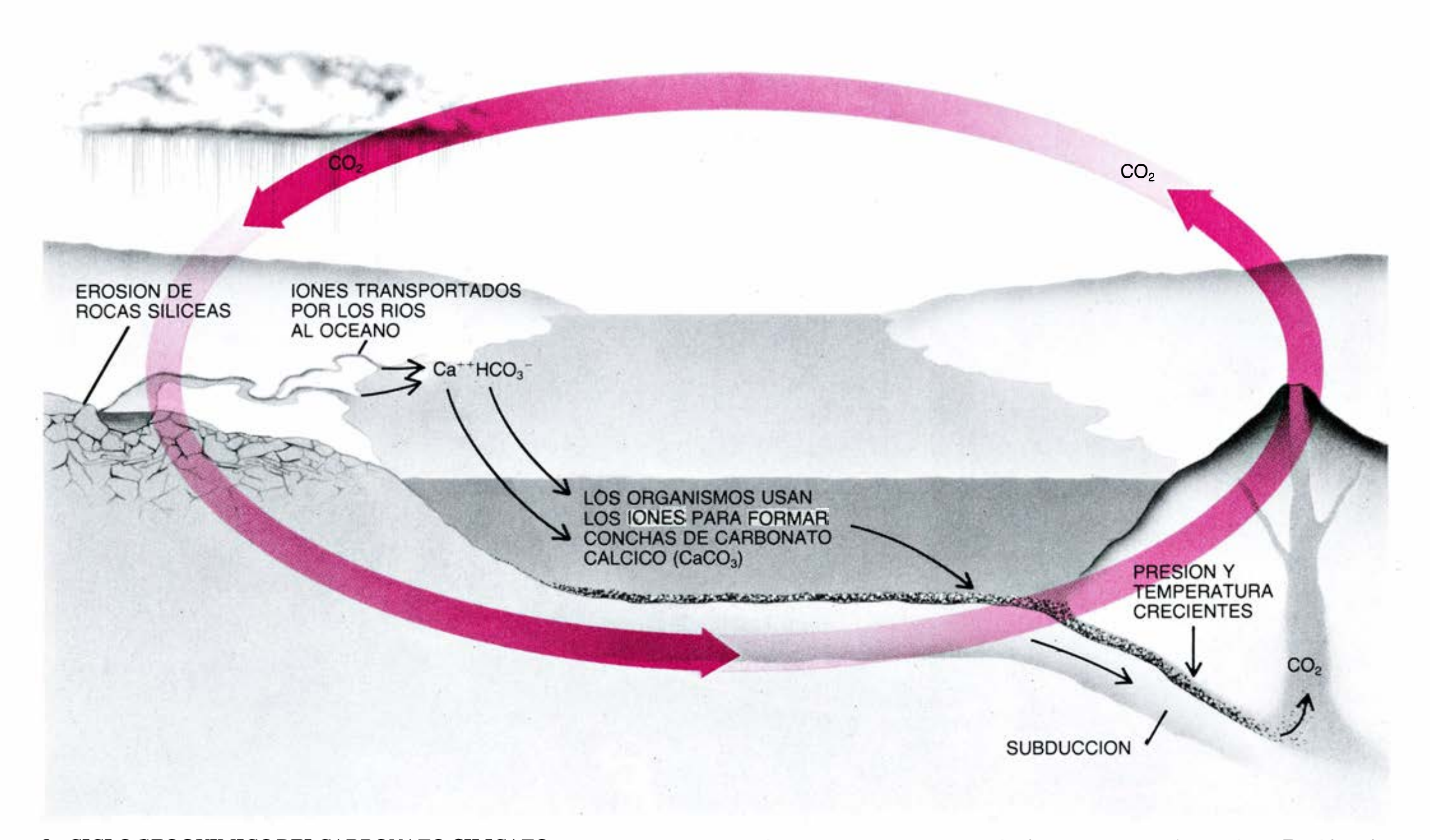
El papel de los organismos

Los organismos desempeñan un papel destacado en el intercambio de dióxido de carbono con la atmósfera. Apoyados en ello, algunos investigadores han sugerido que los seres vivos serían los primeros responsables de la modulación del clima terrestre. James E. Lovelock, de la Estación Experimental Coombe Mill, en Cornualles, y Lynn Margulis, de la Universidad de Boston, son los principales defensores de este punto de vista, que llaman hipótesis Gea ("Gaia"), en honor de la diosa griega de la Tierra. Estos autores sostienen que la disminución del dióxido de carbono atmosférico en el curso de la historia ha sido consecuencia directa de la intervención biológica y que, sin los organismos, el clima terrestre podría muy bien haber seguido los derroteros de Marte o Venus.

El reino de lo vivo desarrolla una acción importante. La fracción de dióxido de carbono (en torno al 20 por ciento) que no participa en el ciclo del carbonato-silicato se elimina de la atmósfera por la fotosíntesis de las plantas. Cuando tales organismos mueren, depositan carbono orgánico en los sedimentos. El dióxido de carbono se regenera cuando los procesos tectónicos elevan las rocas sedimentarias y forman montañas, permitiendo que el carbono de las rocas reaccione con el oxígeno atmosférico del agua de lluvia.

Los organismos condicionan también el ciclo del carbonato-silicato. Ya se ha explicado el papel del plancton oceánico en la formación de sedimientos carbonatados, pero es posible que las plantas terrestres tengan una función más importante aún. Cuando las plantas se descomponen, la oxidación de sus restos eleva la concentración de dióxido de carbono en el suelo. En consecuencia, las concentraciones de dióxido de carbono en el suelo son hoy probablemente más altas de lo que eran antes de la aparición de las plantas vasculares, hace unos 400 millones de años. Esta elevación acelera la conversión de silicatos minerales en sedimentos carbonatados.

Expuesto todo lo anterior, creemos, no obstante, que los controles fundamentales de los niveles de dióxido de carbono atmosférico pertenecen al dominio de la física más que al de la bio-



3. CICLO GEOQUIMICO DEL CARBONATO-SILICATO, que opera en una escala de tiempo superior a 500.000 años; elimina dióxido de carbono de la atmósfera, lo almacena en las rocas carbonatadas y después lo devuelve al aire. Los carbonatos se forman cuando el dióxido de carbono se disuelve en el agua de lluvia y reacciona químicamente con rocas que contienen silicatos minerales de calcio (compuestos formados por calcio, silicio y oxígeno). Tales reacciones liberan iones calcio y bicarbonato (Ca^{++} y HCO_3^-) en el agua contenida en el suelo, que transporta, a su vez, los iones a arroyos, ríos y océano. En el interior de éste, el plancton y otros organismos usan los iones para construir caparazones de carbonato cálcico ($CaCO_3$); al morir, depositan las conchas en los sedimentos del fondo del mar. Lentamente, el suelo marino se extiende, se desliza bajo los continentes y se dobla hacia abajo, transportando el sedimento con él. Sometido a temperatura y presión crecientes, el sedimento desprende gas dióxido de carbono, que tornará luego a la atmósfera en las erupciones de los volcanes.

logía. Podríamos argumentar, por ejemplo, que si no existieran los organismos dotados de concha que depositan carbonato cálcico en el fondo del mar, la concentración de iones calcio y bicarbonato en el agua del mar aumentaría. Una vez que las concentraciones de iones alcanzaran cierto nivel crítico, el carbonato cálcico se formaría sin intervención de organismos. Tal debió ocurrir hace más de 600 millones de años, cuando aparecieron los organismos portadores de caparazón.

De manera análoga, los cálculos muestran que la caída en la descomposición de silicatos instada por la desaparición de las plantas terrestres podría quedar compensada con un aumento de temperatura de unos 10 grados, cambio que podría surgir del bucle de realimentación negativa del ciclo del carbonato-silicato. El aumento del efecto de invernadero produciría un clima semejante al de hace 100 millones de años, durante el Cretácico medio: cálido, aunque adecuado para muchas formas de vida, dinosaurios incluidos. De aquí que haya buenas razones para creer que la Tierra hubiera permanecido habitable, aun cuando nunca hubiera estado habitada. El ciclo del carbonato-silicato, por sí solo, habría producido el necesario mecanismo amortiguador.

Cabría preguntar si el vapor de agua, responsable de buena parte de los 35 grados Celsius del efecto invernadero terrestre, fue responsable también de que el planeta se mantuviera templado en el curso de su evolución. La respuesta es negativa. La cantidad de agua de la atmósfera no contrarresta las variaciones de la temperatura en superficie, sino que las intensifica. La abundancia de agua en la atmósfera aumenta cuando lo hace la temperatura en superficie; disminuye cuando baja dicha temperatura. De ello se infiere, pues, que sólo una disminución global de los niveles de dióxido de carbono podría explicar el fenómeno siguiente: a medida que el Sol fue adquiriendo mayor luminosidad, no se registró ningún aumento paralelo en la temperatura de superficie de la Tierra, sino que se mantuvo dentro de unos límites habitables.

La compensación falla en Marte

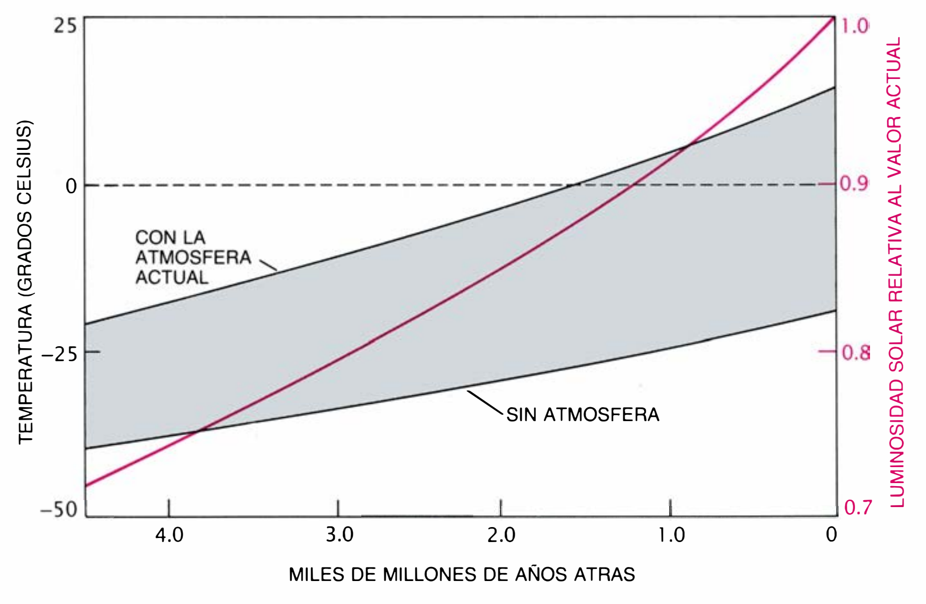
El ciclo del dióxido de carbono puede haber mantenido el clima terrestre dentro de límites razonables mientras el planeta evolucionaba. Si en Marte hubo un proceso similar, fue incapaz de hacer lo mismo en ese planeta. La actual atmósfera está allí enteramente constituida por 0,006 bar de dióxido de carbono, cuyo efecto de invernadero produce un calentamiento de sólo unos 6 grados Celsius.

¿Cabe la posibilidad de que Marte fuera frío desde el principio y que su clima haya experimentado pocos cambios en los últimos 4600 millones de años? No parece verosímil. Las fotografías tomadas por los vehículos espaciales Mariner y Viking, de la NASA, muestran que la superficie marciana está surcada por muchos canales que fueron, casi con certeza, excavados por agua corriente. Aunque algunos canales pudieron haberse formado en un clima frío por brusca liberación de agua desde grandes profundidades, se cree que las redes de valles que surcan los terrenos más antiguos de Marte requirieron, para su formación, temperaturas más altas. Durante los primeros mil millones de años de su historia, el planeta presentó también una tasa de erosión más alta que la de ahora, según estima Peter H. Schultz, de la Universidad de Brown. Este descubrimiento sugiere también que Marte conoció, tiempo atrás, temperaturas suficientes para retener agua líquida.

Los geólogos ignoran cuál fue el margen de temperaturas del planeta, pero su superficie pudo muy bien permanecer cálida bajo el efecto de invernadero de una densa atmósfera de dióxido de carbono. De acuerdo con nuestros cálculos, una atmósfera de entre uno y cinco bares de dióxido de carbono habría mantenido algunas partes de la superficie de Marte por encima de cero grados Celsius en la historia primitiva del planeta. La cifra más baja se aplica a las condiciones en el ecuador marciano, cuando el planeta se halla en la máxima proximidad del Sol; la más alta es un promedio que se aplica al planeta entero.

Entra dentro de lo posible que Marte tuviera alguna vez todo ese dióxido de carbono, aun cuando esas cantidades representan de 150 a 800 veces más del que existe en su atmósfera actual. Si Marte, cuya masa viene a ser la décima parte de la terrestre, alcanzó a poseer una cantidad proporcional de dióxido de carbono, el planeta habría tenido una dotación total de dióxido de carbono equivalente a unos 10 bares. (Para llegar a esta cifra, hay que tomar en consideración la menor superficie y gravedad en superficie de Marte, comparadas con las de la Tierra.)

Ello nos induce a pensar que Marte tenía una proporción adecuada de dióxido de carbono, pero se enfrió porque su mecanismo de reciclado se fue a pique. Creemos que el planeta conoció, tiempo atrás, un sistema eficaz de reciclado; pues, sin él, la erosión de las rocas hubiera eliminado todo el dióxido de carbono atmosférico en unos 10 millones de años. Ahora bien, todo indica que la atmósfera retuvo abundantes cantidades del gas durante un tiem-



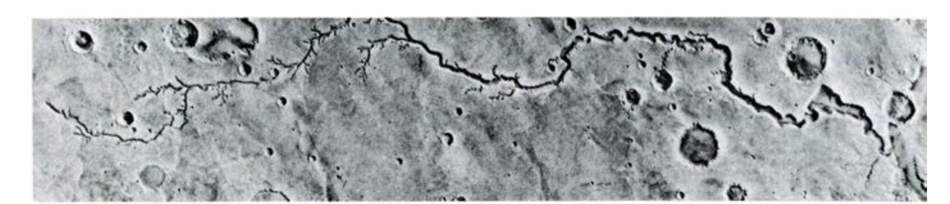
4. LOS CALCULOS obtenidos mediante modelos climáticos indican que la Tierra se habría congelado durante la primera parte de su historia si su composición atmosférica hubiera sido la misma de hoy. La razón estriba en que el Sol era hasta un 30 por ciento más débil en el pasado (*curva en color*). La curva superior muestra la temperatura en superficie calculada mediante un modelo climático "monodimensional" (promediado globalmente), suponiendo una concentración constante de dióxido de carbono. (Muchos de los cálculos explicados en el texto del artículo se basan también en un modelo monodimensional.) La curva inferior representa la temperatura en la superficie de una Tierra sin atmósfera. La región sombreada entre ambas curvas representa la magnitud del efecto de invernadero. Los niveles de dióxido de carbono eran, quizá, más altos en el pasado; la temperatura superficial de la Tierra era también mayor que la mostrada aquí. La curva de luminosidad solar se basa en un cálculo de Douglas O. Gough, de Cambridge.

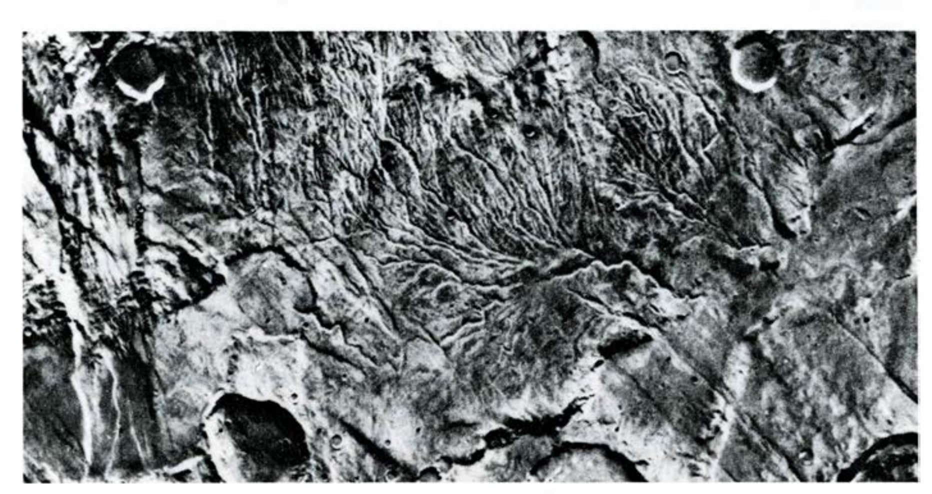
po bastante más largo. Las redes de valles suministran la clave: las tendidas por las antiguas tierras altas del Sur se pueden datar contando el número de cráteres meteoríticos superpuestos. Apoyados en ello, los investigadores han llegado a la conclusión de que las redes persistieron hasta casi el final del período de máximo bombardeo meteorítico, es decir, hace unos 3800 millones de años.

El sistema de reciclado eliminaba probablemente el dióxido de carbono de la atmósfera por los mismos procesos de erosión que en la Tierra. El mecanismo que devolvía el gas a la atmósfera podía haber diferido bastante, sin embargo, porque un planeta tan pequeño como Marte quizá nunca ha desarrollado tectónica de placas. Cabe la posibilidad de que la lava expulsada por los volcanes en la superficie de Marte cubriera sedimentos carbonatados, sepultándolos gradualmente hasta una profundidad donde la presión y la temperatura los obligasen a desprender dióxido de carbono gaseoso. Los modelos de ordenador indican que tal proceso podría haber sido suficiente para reciclar los carbonatos hasta unos mil millones de años después de la formación del planeta.

Al parecer, Marte no se enfrió porque recibiese menos energía solar que la Tierra, sino por su menor tamaño. Poseía también menos calor interno cuando se formó, y su alta relación de superficie a volumen le obligó a perder ese calor a mayor velocidad. Con el tiempo, el interior se hizo tan frío que no pudo ya liberar dióxido de carbono a partir de las rocas carbonatadas. Todo el dióxido de carbono que se eliminara de la atmósfera mediante la erosión quedó retenido en el suelo. La atmósfera marciana se adelgazó y el clima se fue acercando a su presente estado gélido. De haber alcanzado el tamaño de la Tierra, Marte hubiera gozado de grandes probabilidades de conservar suficiente calor interno para continuar reciclando el dióxido de carbono y compensar así el bajo nivel de energía solar recibida.

De acuerdo con este supuesto, Marte encerraría considerables cantidades de rocas carbonatadas enterradas en su corteza. Hasta ahora, las exploraciones espectroscópicas realizadas desde la Tierra no han logrado descubrir tales materiales. Por otra parte, James L. Gooding ha detectado recientemente pequeñas cantidades de carbonato cálcico en los llamados meteoritos SNC (Shergotty, Nakhala y Chassigny), que son fragmentos de roca originados, al parecer, en la superficie marciana. La próxima misión Observador Marciano, proyectada para 1992, llevará a cabo una búsqueda más completa de carbonatos y proporcionará posiblemente nuevas e importantes pruebas referentes a nuestra teoría acerca de la congelación de Marte.





5. SUPERFICIE DE MARTE surcada por muchos canales. De ello se infiere que el planeta alcanzó, tiempo atrás, una temperatura suficiente para conservar líquida el agua. Los canales de escorrentía típicos, como el Nirgal Vallis (*arriba*), ofrecen una morfología distinta de los ríos de la Tierra porque sus tributarios son cortos y escasos; proceden del excavado o afloramiento del agua del suelo. Otras estructuras ramificadas (*abajo*), que se encuentran en suelos antiguos, parecen ser redes de valles; la excavación o la precipitación y ulterior escorrentía pudo haberlas formado. El hecho de que las redes presentan cráteres superpuestos implica que se formaron antes del final del período de intenso bombardeo meteorítico, hace unos 3800 millones de años. Canales "aliviaderos" pudieron haberse formado en un clima frío mediante otros procesos.

Cómo se secó Venus

Marte conserva una ingente provisión de agua, aunque congelada; Venus, en cambio, está casi por entero deshidratado. La escasa cantidad que allí queda reside en la atmósfera en forma de vapor, o integrada en las densas nubes de ácido sulfúrico que rodean el planeta. Los climatólogos han propuesto dos teorías para explicar la sequedad de Venus.

John S. Lewis y sus colaboradores, de la Universidad de Arizona, han sugerido que Venus nunca tuvo mucha agua, esto es, que la región de la nebulosa solar donde se formó Venus era demasiado caliente para permitir la formación de minerales hidratados. Teoría ésta a la que puede planteársele una seria objeción: no tiene en cuenta el papel de la gravedad. De acuerdo con los modelos dinámicos elaborados por George W. Wetherill, de la Institución Carnegie de Washington, los planetas en formación, además de acumular los planetésimos que cruzan sus órbitas, perturban también las órbitas de tales cuerpos, dispersándolos por todo el sistema solar interno. Durante las últimas fases de su crecimiento, los protoplanetas Tierra y Venus tenían masas de tamaño suficiente para haber intercambiado planetésimos. Como los procedentes de la Tierra habrían sido ricos en agua, Venus habría recibido una dotación substancial de ella.

La objeción nos lleva de la mano a la otra teoría: Venus tenía originariamente agua en abundancia, tal vez tanta como la Tierra, pero la perdió cuando llegó hasta la alta atmósfera. Allí, la radiación solar destrozó las moléculas de agua, liberando átomos de hidrógeno que escaparon al espacio. (Sólo el agua de la alta atmósfera está sometida al escape de hidrógeno; a altitudes menores, los átomos de hidrógeno, que son ligeros, se conservan en la atmósfera bajo la influencia de los gases de fondo, como el dióxido de carbono.)

Esta segunda teoría ofrece distintas variaciones según permita o no que el agua permanezca fluida en la superficie durante un tiempo apreciable. La explicación clásica, la del efecto de invernadero descontrolado, sostiene que Venus nunca retuvo agua en su superficie. La noción de efecto invernadero sin control fue ya sugerida en 1955 por Fred Hoyle, de la Universidad de Cambridge, pero debemos muchos de sus

pormenores a Andrew P. Ingersoll, del Instituto de Tecnología de California, y a uno de los autores (Pollack).

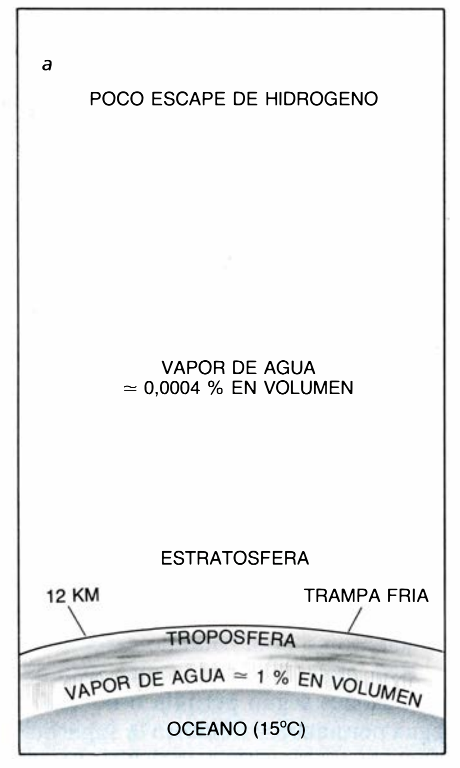
Según dichos investigadores, el agua de la superficie no puede mantenerse en estado líquido si, sobre un planeta, incide energía solar superior a un valor crítico. Si el flujo solar en la órbita de Venus excedió ese valor crítico desde el principio, el agua que hubiera emergido del interior se hubiera vaporizado al instante. Al menos en la parte más baja y cálida de la atmósfera, este vapor no se habría condensado en forma de lluvia y, por tanto, no se habrían formado océanos.

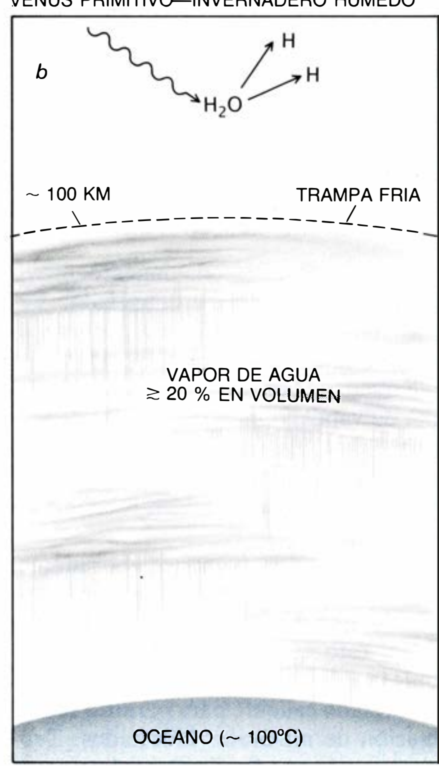
La atmósfera habría perdido agua porque, en un ambiente tan cálido y húmedo, el aire se enfriaría de forma extraordinariamente lenta al ascender. En consecuencia, la "trampa fría" atmosférica habría subido hasta una altitud muy elevada (unos 100 kilómetros). La trampa fría es la región donde la baja temperatura y la alta presión ambiental se combinan para mantener el punto de saturación en un valor mínimo. Normalmente, la concentración relativa de vapor de agua (la fracción de volumen atmosférico representada por el vapor) es mucho más baja en la trampa fría que en la atmósfera subyacente; el agua se condensa en vez de ascender. Pero en una trampa fría a gran altura, la concentración relativa de vapor de agua sería semejante a la que hay en la atmósfera mucho más cerca de la superficie. Bajo estas condiciones, la trampa fría permitiría que una proporción apreciable de vapor de agua pasara a las partes altas, donde quedaría sometida a la fotodisociación y al escape de hidrógeno. Tal escape podría haber eliminado el equivalente a un océano en menos de 30 millones de años.

En la presente atmósfera terrestre, por el contrario, la trampa fría se encuentra a alturas relativamente bajas (entre nueve y 17 kilómetros), en la frontera entre la troposfera y la estratosfera (frontera llamada tropopausa). Cuando el vapor de agua procedente de alturas menores asciende hasta la trampa fría, se condensa; ello explica que nuestra estratosfera sea muy seca y escape poco hidrógeno.

Calculamos que el flujo solar necesario para desencadenar un efecto de invernadero descontrolado viene a ser de 1,4 veces la cantidad de energía solar que incide actualmente sobre la Tierra, siempre que el planeta en cuestión tenga una atmósfera saturada y libre de nubes. Esta cantidad de energía equivale al flujo solar estimado en la órbita de Venus en los albores de la historia del sistema solar, lo que sugiere que Venus estaba al borde de un efecto de invernadero descontrolado. Ello no obstante, si hubiera habido nubes capaces de reflejar una proporción substancial de la radiación solar incidente, es probable que se hubiera evitado un efecto de invernadero descontrolado en el Venus primigenio, permitiendo la existencia de océanos durante algún tiempo.

Tales océanos no se hubieran salvado indefinidamente. Como hipótesis alternativa a la teoría del efecto de invernadero descontrolado, proponemos que Venus tuvo océanos, pero los perdió a causa de un "invernadero húmedo"; entendemos por tal una condición en la que la concentración relativa de vapor de agua cerca del suelo da cuenta de más del 20 por ciento del volumen. Para una atmósfera de un bar, como la de la Tierra, esta concentración se alcanza cuando la temperatura en superficie sube por encima de 70 grados C. (Si Venus hubiera tenido un océano y







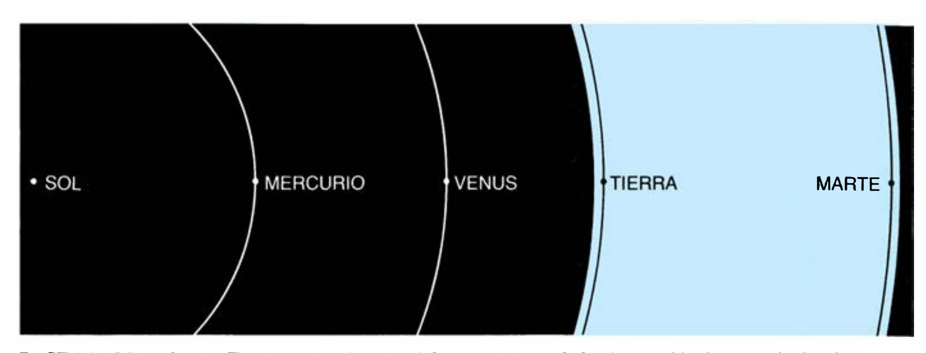
6. TENDENCIA DEL VAPOR DE AGUA a escapar de la Tierra: es mínima. No puede decirse lo mismo del Venus primitivo. En la Tierra (*a*), el agua de la troposfera no puede penetrar en la estratosfera a causa de la trampa fría, la región donde la baja temperatura y la presión atmosférica relativamente alta se combinan para minimizar la concentración de vapor de agua. Cuando el vapor llega a la trampa, se condensa en su mayor parte. En el primitivo Venus, la baja atmósfera, aunque caliente en comparación con la de la Tierra, pudo haber estado suficientemente fría para que el vapor se condensara y formase un océano. Pero este habría desaparecido por culpa del "invernadero húmedo" (*b*): situación que se produce cuando, a causa de una alta temperatura superficial, el vapor de agua constituye más del 20 por ciento de la baja atmósfera. La trampa fría se traslada entonces a mayores alturas y se torna ineficaz para impedir que el vapor de agua llegue hasta la alta atmósfera. Aunque parte del vapor se condensa en forma de lluvia, el que no lo hace se disocia en las alturas y los átomos de hidrógeno que lo forman escapan al espacio. Venus pudo haber alcanzado una temperatura tal que, en vez de lo anterior, se produjera un efecto invernadero descontrolado (*c*): el agua liberada por el planeta se convirtió en vapor y no se formó ningún océano. El agua sólo viajó en una dirección: hacia arriba.

lluvia, la mayor parte de su dióxido de carbono habría quedado enterrado en las rocas carbonatadas y hubiera sido posible una atmósfera de un bar.)

Nuestras simulaciones climáticas indican que el efecto de invernadero húmedo debe producirse cuando la radiación solar incidente sobre una atmósfera libre de nubes sea al menos 1,1 veces la cantidad de radiación solar incidente sobre la Tierra. Cuando la concentración del vapor de agua en la proximidad del suelo sobrepasa el 20 por ciento, la condensación del agua (que genera calor) calienta bastante la atmósfera y, como en el caso del efecto de invernadero descontrolado, eleva la trampa fría. El agua puede llegar entonces hasta la alta atmósfera. Podría existir océano en un planeta con un flujo solar de entre 1,1 y 1,4 veces el de la Tierra, pero se vaciaría, por escape del hidrógeno, en escasos cientos de millones de años.

Desde nuestro punto de vista, la teoría del invernadero húmedo explica mejor que la del invernadero descontrolado el motivo por el que Venus apenas si tiene agua líquida. Como la erosión reduciría los niveles de dióxido de carbono atmosférico en un invernadero húmedo, la presión atmosférica total sería más baja que en el caso del invernadero descontrolado. En virtud de ello, bastaría una pequeña cantidad de vapor de agua para constituir el 20 por ciento del volumen gaseoso total y, por tanto, una mayor fracción del agua total existente alcanzaría la atmósfera superior. Por ejemplo, si la atmósfera constase de un bar de vapor de agua y un bar de dióxido de carbono, el agua constituiría el 50 por ciento del volumen y mucha de ella escaparía. Pero si hubiera 99 bares de dióxido de carbono, el bar de agua constituiría el 1 por ciento del volumen y permanecería en la atmósfera del planeta.

Con independencia de si la primitiva atmósfera de Venus se hallaba en estado de invernadero descontrolado o de invernadero húmedo, el planeta habría alcanzado con el tiempo su presente condición ardiente y seca. Una vez desaparecidos los océanos, cesaría la formación de carbonatos, provocando la acumulación del dióxido de carbono en la atmósfera. En consecuencia, la atmósfera actual del planeta, de 93 bares, sería, sobre todo, de dióxido de carbono. Los gases sulfurosos, escasos en un comienzo porque se disuelven fácilmente en el agua, se acumularon también y formaron las nubes de ácido sulfúrico que constituyen hoy una de las características más importantes de la atmósfera de Venus.



7. SE LLAMA ZONA DE HABITABILIDAD CONTINUA (*azul claro*) la región del espacio donde un planeta podría mantener, en teoría, un clima semejante al de la Tierra durante un tiempo suficiente para que proliferase la vida. Una primera estima sugirió que la zona era bastante estrecha, extendiéndose desde unas 0,95 unidades astronómicas hasta 1,01 U.A., a ambos lados de la órbita terrestre. Trabajos más recientes sugieren que el borde externo podría llegar hasta 1,5 unidades astronómicas, allende la órbita de Marte.

Es el dióxido de carbono, y no la distancia de Venus al Sol, lo que explica las altas temperaturas que hoy se encuentran en su superficie. Venus recibe 1,9 veces más radiación solar que la Tierra, pero sus nubes de ácido sulfúrico reflejan aproximadamente el 80 por ciento de la misma; Venus absorbe, pues, bastante menos energía solar que la Tierra. Sin el efecto de invernadero, Venus estaría más frío que la Tierra y sólo un poco más caliente que Marte.

La zona de habitabilidad continua

El descubrimiento de que un planeta con 1,1 veces el flujo solar de la Tierra perdería su agua por fotodisociación concuerda con el cálculo de Hart según el cual la frontera interior de la zona de habitabilidad continua se encuentra a unas 0,95 unidades astronómicas. Ahora bien, se trata de una concordancia meramente casual: nosotros basamos nuestros cálculos en las tasas de escape del hidrógeno, mientras que él llegó a su previsión por otros medios. Ni que decir tiene que un planeta instalado en la frontera interna no sería habitable por mucho tiempo. El Sol aumenta actualmente su luminosidad en alrededor de un 1 por ciento cada 100 millones de años, lo que implica que la propia Tierra puede verse en dificultad para conservar su agua dentro de los próximos mil millones de años. Este desastre puede aplazarse cierto tiempo mediante una disminución del dióxido de carbono atmosférico propiciada por el ciclo del carbonato-silicato. Tal disminución, sin embargo, podría ser perjudicial para los organismos, porque muchas plantas serían incapaces de realizar su fotosíntesis si recibieran menos dióxido de carbono del que hoy disponen. (Los lectores podrían recordar, a este propósito, que los niveles de dióxido de carbono aumentan en la atmósfera debido al consumo de combustibles fósiles. Pero ese proceder no puede continuar más allá de unos pocos cientos de años, porque las reservas de carbón y petroleo se agotarían. Tras el breve periodo de calentamiento, el nivel de dióxido de carbono volverá a caer de nuevo.)

El límite exterior de la ZHC debe quedar bastante más lejos de lo imaginado por Hart, tal vez hasta 1,5 unidades astronómicas, lo que colocaría dicho límite algo más allá de la órbita de Marte. Restringimos la frontera exterior a esa distancia, porque parece poco verosímil que un planeta terrestre se formara más lejos.

El mecanismo de realimentación negativa que ha ayudado a estabilizar el clima de la Tierra durante los últimos 4500 millones de años actuaría, en líneas generales, en un planeta de tamaño similar y más lejano al Sol. Marte se congeló por la exclusiva razón de su tamaño, demasiado pequeño para continuar reciclando el dióxido de carbono. De acuerdo con nuestra teoría, un planeta del tamaño de la Tierra situado en la órbita de Marte debería tener varios bares de dióxido de carbono en su atmósfera y una temperatura media en superficie por encima del punto de congelación. Esa atmósfera, irrespirable para los humanos, sería capaz de mantener algunas formas de vida.

Cuando Hart determinó que la ZHC era extraordinariamente estrecha, su conclusión implicaba una probabilidad pequeña de encontrar planetas semejantes a la Tierra alrededor de otras estrellas, aun cuando abundasen los otros sistemas planetarios. Nuestros cálculos señalan la conclusión opuesta. Si existen otros sistemas planetarios, algo muy verosímil, habrá también una buena probabilidad de hallar planetas habitables. Que alguno de ellos esté o no habitado es una cuestión abierta, que no se puede ya descartar apelando al clima, único, de la Tierra. Tal vez en planetas así haya una versión extraterrestre del problema de Goldilocks.

La adaptable zarigüeya

La zarigüeya de Virginia se adapta rápidamente a los cambios ambientales. Parte de su éxito se debe a una eficaz estrategia de reproducción, capaz de ajustar la proporción sexual de la prole

Steven N. Austad

Desde que, en febrero de 1500, el explorador español Vicente Yáñez Pinzón y sus hombres descubrieron en Brasil una zarigüeya hembra con sus pequeños, estos mamíferos del Nuevo Mundo, que tienen recién nacidos embrionarios, han constituido para los europeos un motivo de excesos retóricos. Como es el caso de muchos animales exóticos, la primera descripción de la zarigüeya la consideraba un conjunto de retazos de especies conocidas. Por ejemplo, Richard Eden, en su traducción de 1555 del *De Orbe Novo* de Pedro Mártir de Anglería, describía la zarigüeya como "una bestia monstruosa con el semblante de zorra, cola de tití, orejas de murciélago, manos de ser humano y pies de mono, que lleva consigo sus cachorros en un vientre externo muy similar a una gran bolsa o saco". El capitán de navío John Smith, jefe de la colonia de Jamestown, que transcribió el vocablo algonquiano que significa "bestia blanca" al inglés "opassum" (del que deriva el actual "opossum" que designa, en ese idioma, la zarigüeya), fue algo más comedido e hizo de ella una mezcla de rata, gato y cerdo.

Pero la familiaridad envilece. Hoy, la zarigüeya –uno de los mamíferos suburbanos más comunes en América del Norte– despierta poco respeto. Por ejemplo, el zoólogo Durward L. Allen califica la zarigüeya de "perezosa, maloliente e infame criatura, sin traza alguna de personalidad o de dignidad". Además, por ser marsupial, se la considera primitiva y, por consiguiente, inferior a los mamíferos euterios o placentarios. En relación a su tamaño corporal, los marsupiales tienen la temperatura del cuerpo más baja, la tasa metabólica inferior y el cerebro más pequeño que los euterios. Presentan asimismo menos cromosomas, por término medio, además de una cáscara membranosa en torno a los óvulos y, excepción hecha de los peramélidos, una placenta rudimentaria.

Sin embargo, cuesta sostener el argumento de la inferioridad de los marsupiales, sobre todo en lo que se refiere a las zarigüeyas de América. En realidad, el de la zarigüeya norteamericana, o de Virginia, (*Didelphis virginiana*), es uno de los casos modernos de éxito ecológico. La especie colonizó la mayor parte de América del Norte durante el Pleistoceno tardío y ha proseguido su expansión en épocas históricas. Desde la época colonial, ha ampliado su área de distribución en el este de Estados Unidos unos 800 kilómetros hacia el norte.

Este mismo siglo, algunos animales han conseguido que la especie esté representada en la costa oeste, donde se la desconocía con anterioridad a las postrimerías del siglo XIX. En 1906, cerca de Los Angeles, un trampero anónimo capturó la primera zarigüeya silvestre al oeste de las Montañas Rocosas. Se ignora si el ejemplar descendía de alguno de los que trajo de Missouri 20 años antes un tal "tío Billy" Rubottom, o si algún otro nostálgico emigrante había dado suelta a algunas zarigüeyas hasta entonces domesticadas. En 1927, un zoólogo escribió: "La zarigüeya es tan prolífica y tiene una distribución tan amplia en California que ya es demasiado tarde para exterminar esa peligrosa especie". Una década después medraba desde Baja California hasta la frontera canadiense.

Buena parte del éxito de la zarigüeya de Virginia proviene, sin duda, del mutualismo que ha establecido con el ser humano. El hombre le brinda establos, portales y una gama entera de potenciales madrigueras, a la par que una provisión constante de basura comestible. Por su parte, la zarigüeya se erige en un recurso alimenticio, si no epicúreo, al menos fiable. Las excavaciones arqueológicas indican claramente que la zarigüeya constituía un componente habitual de la dieta de los indios de ambos hemisferios americanos. Por ejemplo, las zarigüeyas que viven hoy en diversas islas del Caribe carecen de testimonio paleontológico anterior a la colonización humana. Es probable, por tanto, que se introdujeran como provisión de carne.

El área de distribución de la zarigüeya norteamericana no es mero artefacto de la expansión territorial del ser humano. Recientes investigaciones demuestran la extraordinaria capacidad de adaptación de ese animal. Por ejemplo, a pesar de su pequeño cerebro y su insulsez, tiene un notable talento para encontrar alimento y recordar dónde lo halló. Puesta a prueba su aptitud para recordar, entre cuatro pistas, cuál llevaba a una caja con comida, la puntuación de la zarigüeya superó la de los gatos, pollos, perros, cabras, cerdos, conejos, ratas y tortugas, si bien era inferior a la del ser humano.

No sorprende menos la fuerte resistencia de la especie al veneno de los ofidios de la subfamilia de la serpiente de cascabel –los crotalinos–, en la que figuran también la surucucú, la labaria, la cabeza de cobre y el mocasín acuático. Tal inmunidad parece más una respuesta adaptativa específica a una amenaza ambiental que un subproducto accidental de su fisiología, pues la zarigüeya no es inmune a la mordedura de la mayoría de las serpientes del Viejo Mundo, entre ellas la cobra y la víbora bufadora. Sin embargo, tal inmunidad frente a los crotalinos no sólo le permite escapar de un potencial depredador, sino también comérselos impunemente.

En nuestras propias investigaciones hemos descubierto que la zarigüeya ajusta la proporción sexual de la progenie para alcanzar un éxito reproductor óptimo con cantidades variables de inversión parental. Esta prueba de que la zarigüeya alcanza un elevado rendimiento reproductor en condiciones ecológicas diversas quizás ayude a explicar cómo desarrolló la amplia serie de complejas adaptaciones que le per-

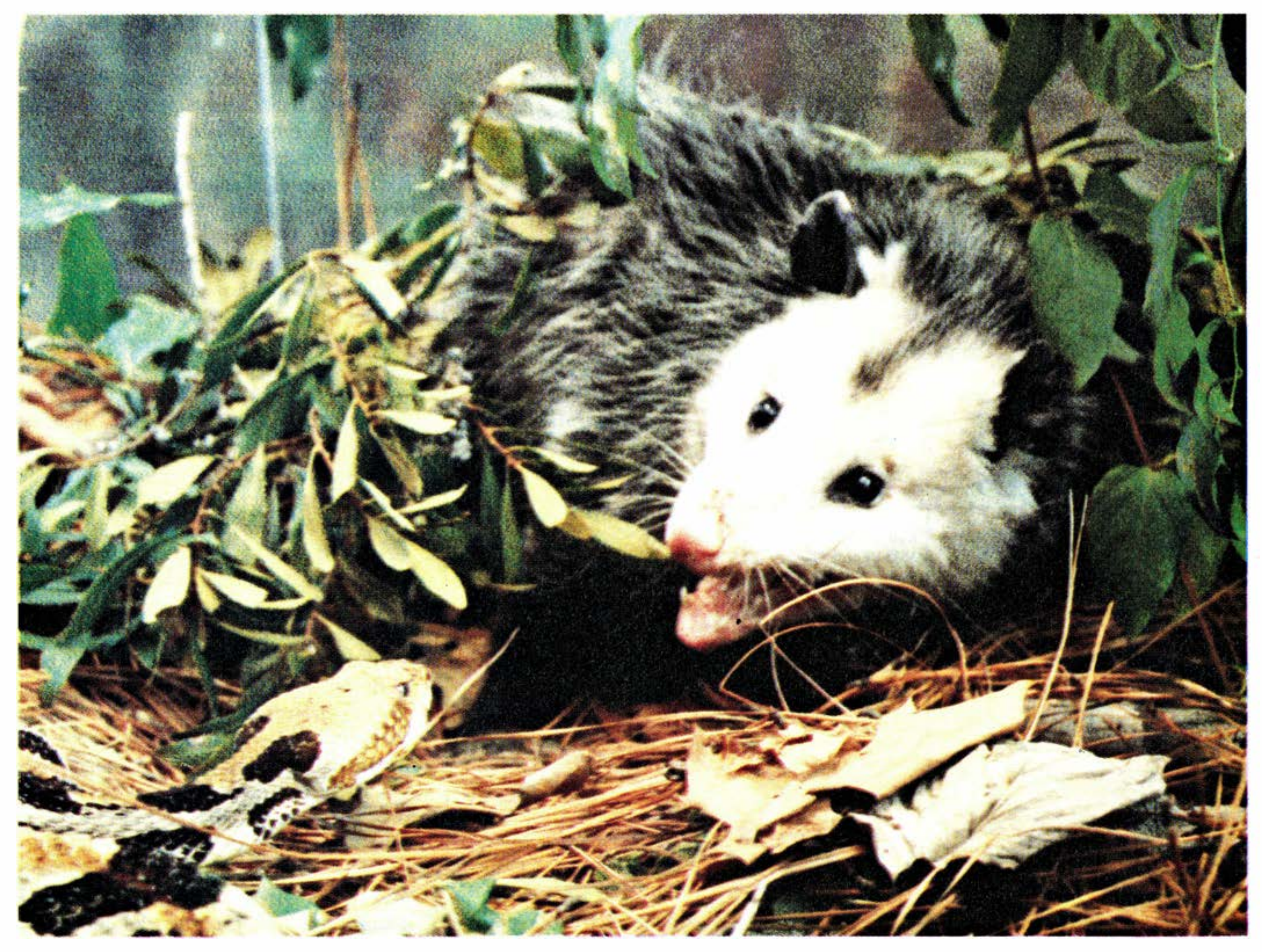
mitieran competir, con éxito, con sus parientes mamíferos más "avanzados".

Los marsupiales y los euterios divergieron hace unos 100 millones de años. Hace unos 35 millones, Norteamérica poseía una rica fauna marsupial, que incluía no menos de tres familias, cinco géneros y trece especies. Los fósiles más antiguos encuadrables en el grupo de los marsupiales proceden de Canadá, y la opinión mayoritaria (aunque no unánime) es que se originaron en América, posiblemente en la del Norte. Por razones desconocidas, todos los marsupiales de Norteamérica se extinguieron hace unos 15 millones de años, cuando se estaba difundiendo con éxito por Sudamérica, Australia y Nueva Guinea. Hoy, América Central y del Sur suman, en conjunto, casi 80 especies de marsupiales, lo que viene a representar más del 30 por ciento del total mundial. La familia de las zarigüeyas americanas, los didélfidos, da cuenta por sí sola de unas 70 especies. (No guarda ningún estrecho parentesco con los falangéridos australianos.)

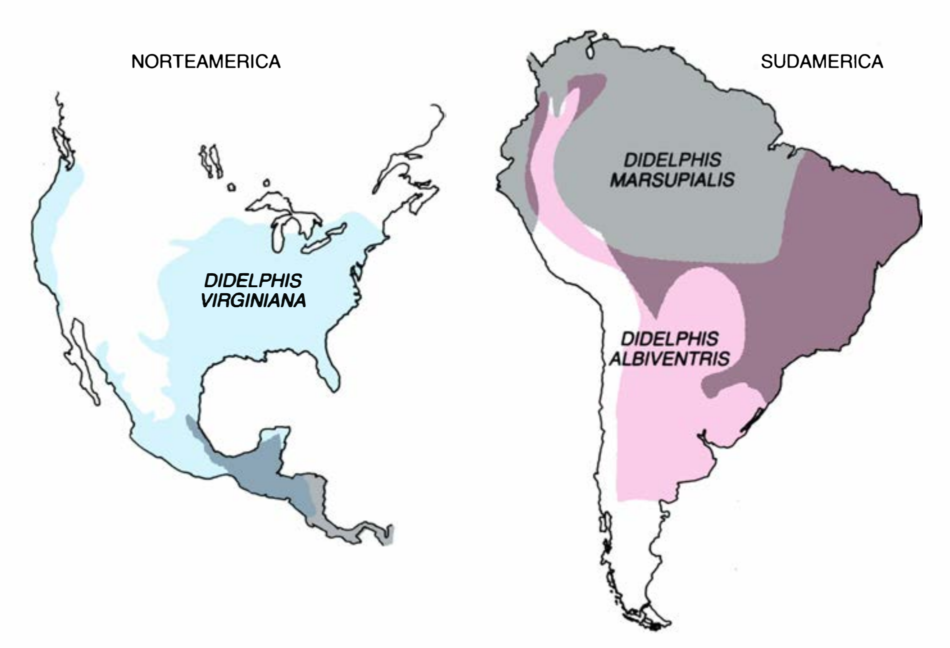
El único marsupial que sobrevive en Norteamérica es la zarigüeya de Virginia. Se trata de una de las tres especies del género *Didelphis*, que se distribuye desde el sur de Canadá hasta las tierras australes de Argentina; ello le convierte en uno de los géneros mamíferos de más amplia distribución del hemisferio y, con mucho, en el marsupial más difundido del mundo. En virtud de la semejanza morfológica que presentan las zarigüeyas *Didelphis* con algunos de los marsupiales fósiles más primitivos, se las ha incluido, erróneamente, entre los fósiles vivos. Mas, por su conservadora morfología, suele recurrirse a la zarigüeya de Virginia como paradigma de la anatomía funcional de los mamíferos primitivos.

A pesar de la apariencia conservadora de su morfología, *Didelphis* constituye un género de evolución reciente y, en algunos aspectos, muy especializado. Los primeros fósiles declaradamente *Didelphis* surgieron hace unos 4 millones de años en Suramérica. Alfred L. Gardner, del Museo Nacional de Historia Natural, sostiene, a partir del análisis cromosómico, del registro fósil y de los datos paleoclimatológicos, que la zarigüeya de Virginia es la benjamina de las tres especies del género, separada de la especie ancestral, la zarigüeya cancrívora (*Didelphis marsupialis*), en los últimos 75.000 años.

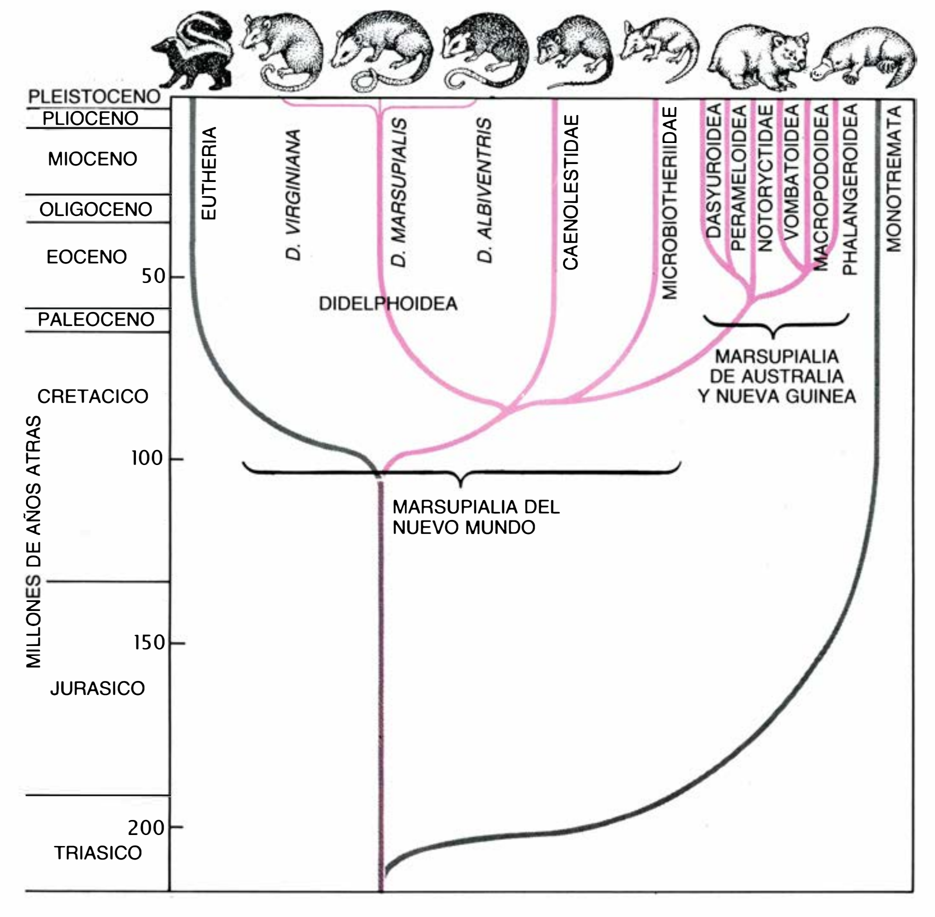
Gardner propone que *D. marsupialis* colonizó México cuando América del Sur y del Norte volvieron a quedar comunicadas por tierra, durante el Plioceno, hace entre cinco y dos millones de años. Sin embargo, como no toleraba ni los climas áridos ni los fríos, no pudo aventurarse en lo que hoy son los Estados Unidos. Durante el Pleistoceno, hace aproximadamente entre un millón y 10.000 años, cuando las repetidas glaciaciones de Norteamérica ocasionaban grandes fluctuaciones climáticas en México, *D. marsupialis* se propagó por las montañas del occidente mexicano durante los períodos cálidos, retirándose a los valles bajos, aislados, durante los fríos. En una de esas pe-



1. ZARIGÜEYA enfrentándose a un crótalo. El marsupial es inmune al veneno de los ofidios de la subfamilia de los crótalos que, a sus ojos, constituyen una fuente de alimento. En ciertas zonas de Texas, representan hasta el 6 por ciento de la dieta de la zarigüeya. La resistencia al veneno podría haberse desarrollado en respuesta adaptativa específica a una amenaza ambiental: la especie, inmune al veneno de las serpientes del Nuevo Mundo, sucumbe al de las del Viejo.



2. AREA DE DISTRIBUCION de las zarigüeyas *Didelphis*, que se extiende desde el sur de Canadá hasta el sur de Argentina, lo que convierte a ese género marsupial en el de más amplia distribución del mundo. La comadreja overa (*D. albiventris*) vive preferentemente en los climas más fríos de Sudamérica. La zarigüeya cancrívora (*D. marsupialis*) habita en las zonas forestales tropicales, cálidas. La de Virginia (*D. virginiana*) emigró a América del Norte durante el Pleistoceno. Medra en la mayoría de los hábitats, excluidos los desiertos y las montañas que superan los 3000 metros de altura. Desde la época colonial, su área de distribución por el continente norteamericano se ha ampliado 800 kilómetros hacia el norte. Desconocida antes allí, los colonos introdujeron la especie en la costa oeste de los Estados Unidos a comienzos de nuestro siglo.

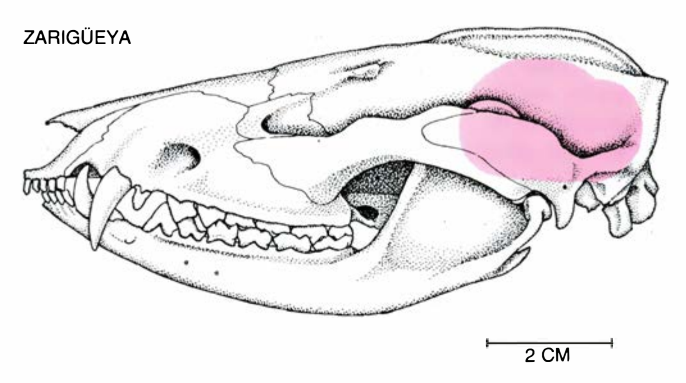


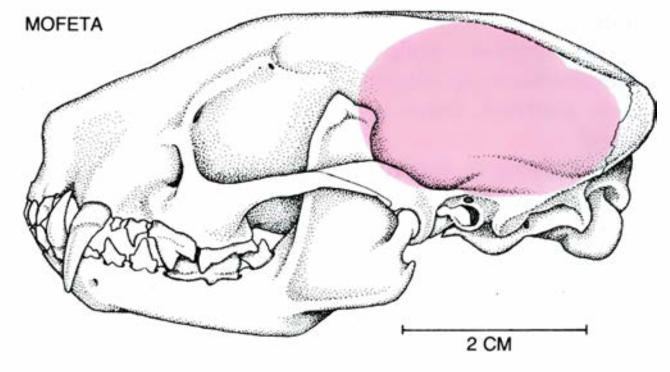
3. ORIGEN DE LOS MARSUPIALES. Parece que tuvo lugar hace alrededor de 100 millones de años, en el Nuevo Mundo. La superfamilia Didelphoidea, a la que pertenece la zarigüeya de Virginia, está considerada la más antigua de las que han sobrevivido hasta el presente. Morfológicamente, *Didelphis* guarda gran semejanza con los fósiles marsupiales más primitivos. *D. virginiana*, que apareció hace tan sólo 75.000 años, es la única especie marsupial que ha conseguido difundirse por la Norteamérica alejada del trópico.

queñas poblaciones aisladas se produjeron importantes reordenaciones cromosómicas, que incomunicaron genéticamente al grupo y, de modo subsidiario, quizá contribuyeron a adaptar la especie incipiente a climas más fríos, lo que le permitiría avanzar hacia el norte. Hace 4000 años, la zarigüeya de Virginia se había difundido por la mayor parte del este de Estados Unidos. Gardner encuentra asimismo pruebas de una endogamia extrema, lo cual podría predisponer a la especie a ser una buena colonizadora, porque la tolerancia a la endogamia facilitaría que con unos pocos animales se fundaran nuevas poblaciones.

Otra característica que contribuiría a la plasticidad ecológica de la zarigüeya es su elevada eficacia reproductora, que comienza antes incluso de la concepción con un empleo inusualmente eficiente del esperma. El macho insemina a las hembras con sólo unos tres millones de espermatozoides, un 5 por ciento de los cuales alcanzan el punto de fecundación; a título comparativo, el conejo insemina a la hembra con unos 150 millones de espermatozoides, de los que el 0,01 por ciento alcanzan el punto de fecundación. La eficiencia de la zarigüeya podría estar vinculada al emparejamiento de los espermatozoides: una característica de los marsupiales americanos que quizá resulte única entre los vertebrados. En *Didelphis*, los espermatozoides se conjugan lateralmente por parejas mientras se hallan aún en los testículos, y nadan juntos por el tracto reproductor femenino antes de separarse en el oviducto. Algunos investigadores creen que el emparejamiento favorece un transporte excepcionalmente bueno, así como una elevada supervivencia en el tracto reproductor. Después de la concepción, las hembras marsupiales invierten mucho menos tiempo en la gestación de las crías que los mamíferos euterios. Los marsupiales recién nacidos son poco menos que embriones móviles, y la mayor parte de su desarrollo transcurre externamente, a menudo en una bolsa. Las zarigüeyas *Didelphis*, por ejemplo, gestan sus crías a lo largo de unos 13 días; al nacer, pesan menos de 0,2 gramos, una diezmilésima parte del peso corporal de la madre. En comparación, la mofeta —mamífero euterio americano de tamaño comparable— gesta sus crías durante 65 días y los cachorrillos pesan unos 33 gramos en el momento del alumbramiento, más de 150 veces lo que un recién nacido de zarigüeya.

Al nacer, las zarigüeyas son sordas y





4. CEREBRO DE ZARIGÜEYA, de pequeño tamaño comparado con el de una mofeta. Otro carácter primitivo es el elevado número de dientes: ningún euterio presenta más de tres incisivos o molares, mientras que la zarigüeya tiene cinco incisivos en la mandíbula superior, cuatro en la inferior y cuatro molares.

ciegas; las extremidades posteriores y la cola son vestigiales. Sin embargo, los miembros anteriores se desarrollan precozmente y están provistos de uñas efímeras, que facilitan al recién nacido el recorrido hasta la bolsa, una distancia de cuatro o cinco centímetros que cada cual ha de salvar por su cuenta, sin ayuda materna. Avanzan con una especie de movimiento de reptación calificado a veces de *crawl australiano*. Nacen más cachorros de los que pueden criarse con éxito: la zarigüeya de Virginia tiene trece pezones y da a luz unos 22 cachorros; sin embargo, el tamaño de la camada superviviente es, por lo común, de seis a ocho.

Durante los primeros 60 días, las crías de la bolsa permanecen aferradas a las tetillas de la madre, que se han hinchado en su boca, inflexible. Transcurridos 70 días, pueden agarrarse y soltarse de las tetillas y hasta recorrer libres distancias cortas. El destete se produce hacia los 100 días, tras lo cual el contacto entre la madre y los pequeños es escaso. Por tanto, la hembra de zarigüeya invierte unos 112 días en sacar adelante la camada desde la fecundación hasta el destete. La mofeta dedica a ello 121 días, un tiempo comparable, pero mientras que las crías de zarigüeya son completamente independientes tras el destete, las de mofeta permanecen en estrecho contacto con la madre durante un mes o más. La temprana independencia permite a la zarigüeya engendrar dos y, ocasionalmente, tres camadas por año, frente a la única de la mofeta. Por consiguiente, el esfuerzo reproductor anual de la mofeta es superior.

Por ser externo el desarrollo y poderse así seguir de cerca, Melvin E. Sunquist, de la Universidad de Florida, y el autor vieron en la zarigüeya un animal idóneo para estudiar cómo distribuyen los animales sus energías reproductoras. En particular, les interesaba contrastar la teoría según la cual los animales ajustaban la proporción sexual de la progenie a la máxima inversión reproductora de que eran capaces. Esta teoría, intuitivamente satisfactoria, desarrollada a comienzos de los años setenta por Robert L. Trivers y Dan E. Willard, de la Universidad de Harvard, defendía que las hembras de mamíferos que podían efectuar una gran inversión en cada descendiente producían sobre todo machos, en tanto que las sólo capacitadas para una inversión pequeña producirían preferentemente hembras.

Se partía, pues, del siguiente supuesto: la aptitud o eficacia biológica de los descendientes masculinos estaría más influida por la cuantía de la inversión parental que la de los femeninos. Como, por lo general, los mamíferos machos tienen más de una pareja y no participan en el cuidado parental, las hembras no suelen carecer de consorte. Por consiguiente, probablemente se reproducirán casi todas ellas, aunque ninguna engendrará tantos descendientes como los machos de más éxito. En cambio, los machos débiles pueden acabar sin pareja o sin descendencia. Luego, suponiendo que la magnitud de la inversión parental influye en la futura aptitud reproductora del animal, los machos deberían verse más favorecidos por una inversión elevada, y más perjudicados por una baja, que las hembras. De ese modo, las madres capaces de ofrecer una gran dedicación habrían de tener sobre todo machos, mientras que las que invierten un esfuerzo reducido tendrían sobre todo hembras.

Por plausible que parezca, hay razones para dudar de que los mamíferos sean capaces de ajustar la proporción sexual de la progenie. En primer lugar, cuesta imaginar un mecanismo verosímil. El sexo del descendiente lo determina el que óvulo lo fecunde un espermatozoide portador de un cromosoma X o un cromosoma Y, y no hay indicios de que los machos controlen la proporción relativa de esos espermatozoides en función de la hembra con la que den en aparearse. En segundo lugar, los animales domésticos, bien tratados y bien alimentados, habrían de desarrollar, a priori, un esfuerzo reproductor mucho mayor que sus parientes silvestres, y sin embargo la mayoría presenta proporciones sexuales equilibradas. Es cierto, por otro lado, que en ocasiones se han propuesto teorías científicas (la tectónica de placas, por ejemplo) antes de descubrirse los mecanismos subyacentes y, puesto que los animales domésticos han estado sometidos a selección artificial durante largos períodos, resultan poco claras las tendencias que pueda seguir en la actualidad la proporción sexual de su progenie.

Pusimos a prueba la hipótesis, reforzando la alimentación de las zarigüeyas hembras silvestres en la esperanza de crear hembras capaces de dedicar una intensa atención. Dotamos a nuestros animales de collares radioemisores, para localizarlos en su guarida durante el día y colocarles comida cerca de la entrada al atardecer, que consumirían al salir en sus correrías de aprovisionamiento nocturnas. Como control, tomamos un segundo grupo de hembras, también con collares radioemisores, que vivían en el mismo hábitat, pero que no recibían alimento. El grupo de prueba lo formaban 20 individuos y, el de control, 18.

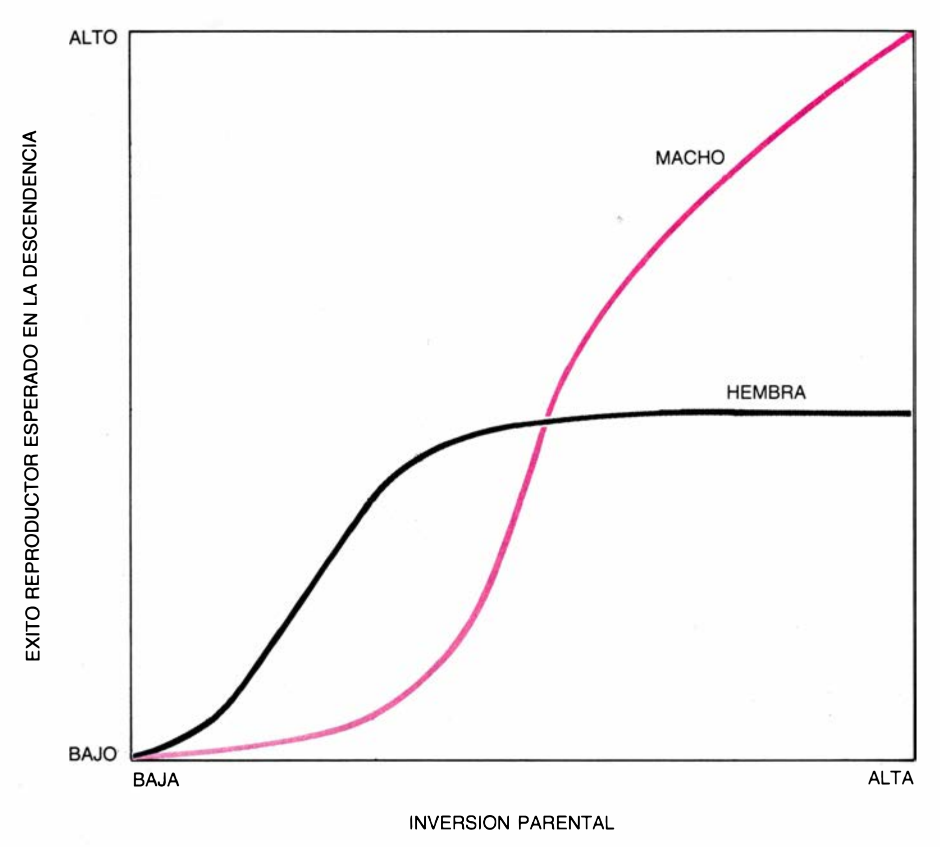
Nuestros resultados fueron concluyentes. Las hembras del grupo de control presentaron una proporción sexual equilibrada; las de alimentación reforzada engendraron más machos. Las hembras de ambos grupos portaban un número equivalente de crías en la bolsa. Sin embargo, las del segundo grupo eran, sistemáticamente, mayores que las del control, lo que denotaba que las hembras alimentadas invertían realmente más en la progenie. Un indicio de que los machos se beneficiaban más de la inversión adicional que las hembras fue que la tasa de recaptura de ju-

veniles era mayor para los machos que para las hembras cuando se comparaba la descendencia de madres con alimentación reforzada frente a la del grupo control.

No realizamos ningún experimento para determinar si las hembras susceptibles de una inversión particularmente reducida engendraban más descendientes femeninos, pero contamos con algunas observaciones que lo sugerían. La mayoría de las zarigüeyas silvestres sólo viven una temporada reproductora. Las pocas que llegan a un segundo año muestran signos de senectud: cataratas, pérdida de peso y falta de coordinación motora. En las colonias de zarigüeyas de Virginia cautivas, las hembras de dos años presentaban asimismo atrofia de los órganos reproductores, disminución del tamaño de la camada y una alta proporción de pérdida de la misma. Esos datos indican que tales hembras quizá no puedan dar tanto como las de un año. Nuestra muestra de hembras de dos años es pequeña, porque no son frecuentes, pero ya hemos conseguido atrapar 19. Las crías de su bolsa presentan un sesgo femenino aún más acentuado que el masculino de las hembras con alimentación reforzada.

Nuestro estudio de las hembras de dos años reveló otra característica insólita: en relación a su tamaño, las zarigüeyas figuran entre los mamíferos de vida más corta. Encontramos que, en la zarigüeya cancrívora, de Venezuela, de un total de 78 hembras maduras con collares radioemisores, el 18 por ciento sobrevivió hasta reproducirse en una segunda estación, el 1,2 por ciento llegó hasta la tercera y ninguna sobrevivió hasta la cuarta. En su investigación de la zarigüeya de Virginia, Sunquist halló que, en Florida, el 8,3 por ciento sobrevivía hasta la segunda estación, el 3 por ciento hasta la tercera y ningún ejemplar llegaba a la cuarta. En colonias de criadero, donde se llevan registros precisos, la duración media de la vida suele ser de dos años y, el máximo, de tres a cuatro años. La zarigüeya se sale también de lo común en otro aspecto: los individuos más viejos muestran signos fisiológicos de senectud —algo que el saber popular considera poco común. Esa creencia popular podría estar equivocada, pues apenas se han realizado estudios de envejecimiento en poblaciones de mamíferos salvajes.

Dos son los enfoques tradicionales del estudio del envejecimiento. El fisiológico trata de determinar qué mecanismos celulares y bioquímicos inmediatos subyacen al propio proceso de envejecimiento. El enfoque evolutivo analiza cómo puede la selección natural dar razón de ese proceso y busca pruebas en las tasas de envejecimiento comparadas de diferentes especies, o de la misma especie en distintos ambientes, así como en las correlaciones genéticas entre la tasa de envejecimiento y otros parámetros del historial biológico, como la tasa de reproducción.

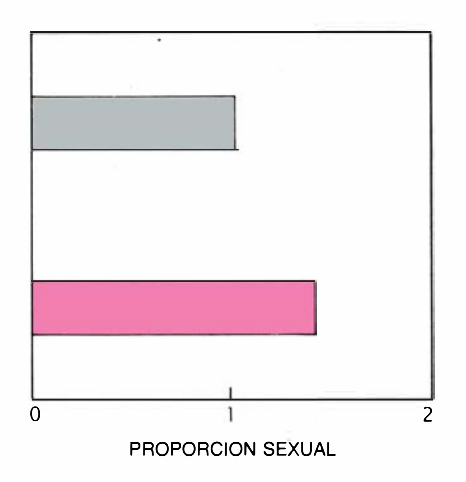


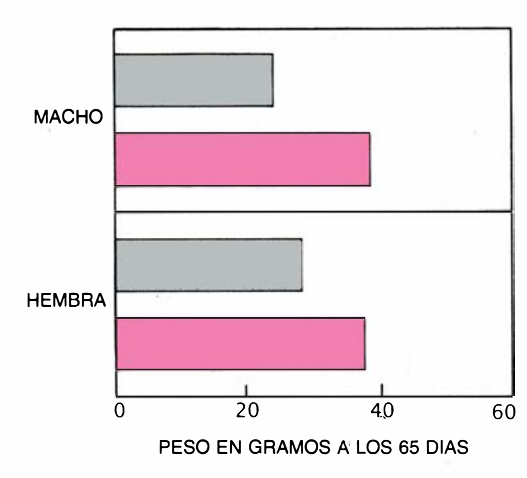
5. HIPOTESIS DE LA PROPORCION SEXUAL, según la cual en los mamíferos poliginos el éxito reproductor varía más ampliamente entre los machos que entre las hembras: los machos fuertes tendrán muchas parejas, pero los débiles quizá no encuentren ninguna, mientras que casi todas las hembras se aparearán. Por consiguiente, si la magnitud de la inversión parental ejerce alguna influencia sobre la futura aptitud reproductora que muestren los descendientes, las madres capaces de realizar una inversión alta deberían producir más hijos machos que hembras y, las menos capaces de ese esfuerzo, más hembras que machos.

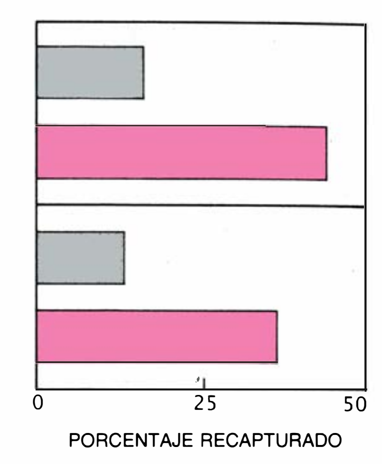
Sir Peter Medawar elaboró, en la década de 1940, la primera teoría evolutiva general del envejecimiento, ampliada en los cincuenta por George C. Williams, de la Universidad estatal de Nueva York en Stony Brook. Medawar señaló que, aunque los animales no se debilitaran progresivamente con la edad, la influencia de la selección natural menguaría según envejecieran, pues crece la probabilidad de muerte accidental —muerte por causas externas al propio animal, como depredadores o enfermedades infecciosas. Por consiguiente, las mutaciones deletéreas que operan al final de la vida sólo acusarán una débil selección en contra. Medawar señaló, asimismo, que la evolución podría favorecer las mutaciones que estimulan la reproducción temprana, aun cuando incrementen la probabilidad de muerte prematura.

Un sencillo ejemplo ilustrará ambos aspectos de la hipótesis. Imagínese una población de individuos que no envejecen. En términos demográficos, ello significa que las probabilidades de muerte y de reproducción con éxito son constantes en el tiempo. Si la probabilidad de muerte accidental es de un 10 por ciento anual, al cabo de 20 años sólo sobrevivirá un 12 por ciento de la población original; a los 40 seguirá con vida poco más de un 1 por ciento. Cualquier mutación deletérea que no se exprese hasta edad avanzada afectará sólo a una pequeña fracción de la población, y toda mutación que provoque la muerte instantánea a los 60 años apenas sufrirá selección en contra. Así, nada impide la acumulación de mutaciones letales o debilitadoras de acción tardía. Si la tasa de muerte accidental es superior, podrán acumularse mutaciones con efectos letales mucho más tempranos.

Según cual fuera la tasa de mortalidad accidental, también podría ofrecer ventaja evolutiva el incremento de la reproducción temprana, aunque conllevara un ineludible coste de supervivencia. Por ejemplo, si la tasa anual de







6. REFUERZO EXPERIMENTAL de la alimentación. Demostró que los descendientes de madres que recibían una alimentación suplementaria (*color*) eran mayores y presentaban una proporción de machos a hembras superior al grupo mantenido como control (*gris*). En comparación con este último, la tasa de recaptura para el grupo de prueba era más alta para los machos que para las hembras, señal de que aquellos sacan más beneficio de una inversión elevada.

mortalidad accidental es del 40 por ciento y cada individuo produce anualmente un descendiente a partir del primer año de edad, entonces cada uno dejará, por término medio, 1,5 descendientes en el curso de su vida. Supóngase ahora que una mutación incrementa la producción a dos descendientes anuales a partir del primer año de edad, pero que, a costa de ese exceso de reproducción temprana, se duplica la probabilidad postreproductora de muerte. Estos individuos producirían, en promedio, algo más de 1,7 descendientes y, por consiguiente, a largo plazo remplazarían a la población original. Una vez más, cuanto mayor es la tasa de muerte accidental, mayor resulta la ventaja de que la reproducción sea temprana e intensa. Luego, en general, se esperará que las poblaciones sujetas a una elevada mortalidad accidental muestren un alto esfuerzo reproductor en edades precoces y que aceleren el envejecimiento; lo contrario será el caso de las poblaciones expuestas a una mortalidad accidental baja.

La principal dificultad que se plantea al verificar en la práctica esa hipótesis en poblaciones naturales es encontrar un parámetro adecuado para el envejecimiento. La longevidad no es una buena medida, porque no se puede distinguir la muerte accidental de la muerte por senescencia. Lo ideal es una medida fisiológica del envejecimiento que pueda aplicarse a cualquier edad y, a un mismo individuo, en diferentes edades. El cambio de las propiedades físicas de las fibras del tendón caulinar de los mamíferos puede constituir un parámetro de ese tipo. El tendón es un tejido simple, formado casi exclusivamente por haces de fibras paralelas de colágeno. A medida que el animal envejece, aumenta de modo regular la resistencia del colágeno y disminuye su solubilidad y flexibilidad. Además, el ritmo de envejecimiento del colágeno de la cola responde a factores, como la desnutrición y los tratamientos hormonales, que, según se sabe, influyen en el envejecimiento de los animales.

Un buen método de medición del envejecimiento del colágeno caulinar —desarrollado a comienzos de los años sesenta por Harry Elden y Robert Boucek, del Instituto Médico Howard Hughes— consiste en registrar el tiempo de ruptura de fibras aisladas, sometidas a cierto peso en una disolución concentrada de urea a una temperatura prefijada. Se ha demostrado que, en los roedores de laboratorio, el tiempo de ruptura aumenta geométricamente con la edad del animal. En el campo, se obtienen haces de fibras de ejemplares capturados mediante un rápido y sencillo procedimiento quirúrgico practicado bajo anestesia local. Acto seguido, se da suelta al animal para que prosiga su vida normal. Muestreando repetidamente el mismo individuo, se calcula la tasa de envejecimiento del colágeno, que admite comparaciones entre animales.

Las dos hipótesis que estamos sometiendo a examen en mi laboratorio de Harvard son: que las zarigüeyas envejecen rápidamente porque dedican un gran esfuerzo a la reproducción precoz y porque están sujetas a cotas excepcionalmente altas de mortalidad accidental. En el curso del ensayo, se reduce de modo artificial el esfuerzo reproductor de las zarigüeyas mediante manipulación del tamaño de las camadas, para comprobar el efecto que esa manipulación pueda ejercer sobre el envejecimiento del colágeno. Puesto que la inversión reproductora prenatal es insignificante, podemos influir de forma notable en el esfuerzo reproductor global sin más que crear un abanico de tamaños de camada, mediante el intercambio de crías entre las hembras. Estudiamos la tasa de muerte accidental comparando el envejecimiento del colágeno, la reproducción y la longevidad de nuestra principal área de estudio, en la Savannah River Plant, Carolina del Sur, con las de las zarigüeyas que viven en Sapelo Island, Georgia, donde falta la mayoría de sus principales depredadores y, por consiguiente, la mortalidad accidental debería ser inferior. El trabajo está en sus fases preliminares. De momento hemos confirmado que, en las zarigüeyas, la técnica del colágeno aprecia diferencias de edad de sólo dos meses, que puede realizarse con éxito el intercambio de camadas y que la población de la isla ha estado incomunicada genéticamente al menos varios miles de años.

Resulta irónico que el éxito ecológico de las zarigüeyas *Didelphis* haya traído consigo, por vía indirecta, el escaso aprecio que suelen merecer. Esa opinión podría cambiar, al menos entre los científicos, en la medida en que el género ayuda a esclarecer diversos aspectos de la biología de los mamíferos. La investigación realizada hasta la fecha indica que las zarigüeyas son animales especialmente idóneos porque se dejan controlar y manipular con facilidad, tanto en el campo como en el laboratorio. Permiten abordar cuestiones ecológicas y evolutivas de ámbito general, a la par que explorar detalles de sus mecanismos. En el futuro, las zarigüeyas podrían revelar los mecanismos mediante los cuales las fuerzas ecológicas determinan la proporción sexual y la tasa de envejecimiento, y los fisiólogos quizás encuentren en ellas animales valiosos para interpretar el proceso del envejecimiento y los mecanismos de resistencia a los venenos.

Plásticos que conducen la electricidad

Los plásticos, baratos, duraderos, ligeros y versátiles, poseen muchas propiedades valiosas, pero la conductividad no es una de ellas. La próxima generación de plásticos acabará con esa tradición

Richard B. Kaner y Alan G. MacDiarmid

Para la mayoría de las personas, el título de este artículo hubiera parecido absurdo hace 20 años, cuando el prejuicio teórico había clasificado a los plásticos entre los aislantes. La sugerencia de que un plástico pudiera ser un conductor tan bueno como el cobre hubiera parecido todavía más ridícula. A pesar de ello, tales maravillas se han logrado durante estos últimos años mediante modificaciones sencillas de los plásticos ordinarios. Los nuevos materiales, denominados polímeros conductores, reúnen las propiedades eléctricas de los metales y las ventajas de los plásticos que despertaron gran expectación en las décadas de 1930 y 1940.

Para conseguir que un polímero conduzca la electricidad, es necesario incorporar pequeñas cantidades de ciertos productos químicos mediante un proceso conocido con el nombre de contaminación o dopado. El proceso de dopado de los polímeros es mucho más sencillo que el utilizado para contaminar el silicio y otros semiconductores ordinarios; de hecho, las técnicas para producir y dopar los polímeros conductores existían ya mucho antes de que la comunidad científica estuviera preparada para admitir la idea de que los plásticos podían ser buenos conductores. Cuando se demostró la posibilidad de que los polímeros condujeran electricidad, la idea conoció un rápido desarrollo. En 1977 se sintetizó el primer polímero conductor; en 1981 se fabricó la primera batería con electrodos de polímeros. En el verano pasado, la conductividad de los polímeros igualó la del cobre y, hace algunos meses, salió al mercado la primera batería recargable fabricada con polímeros.

Los avances subsiguientes avalan la posibilidad de que los polímeros conduzcan la electricidad mejor que el cobre; mejor incluso que cualquier otro material a la temperatura ambiente. Los polímeros pueden incluso reemplazar a los conductores de cobre en aquellos casos en que el peso es un factor importante, como sucede en los aviones. Los polímeros conductores poseen también interesantes propiedades ópticas, mecánicas y químicas; este hecho, junto con su capacidad de conducir la electricidad, puede favorecer su utilización en muchas aplicaciones modernas en las que resulta imposible usar el cobre. Por ejemplo, la aplicación de capas delgadas de estos polímeros en los cristales de las ventanas podría absorber la luz solar; el grado de absorción podría controlarse mediante la aplicación de un potencial eléctrico. Estos polímeros conductores podrían incluso utilizarse para fabricar nervios artificiales susceptibles de implantarse sin rechazo en el cuerpo humano.

El descubrimiento de los polímeros conductores tuvo lugar por accidente. A principios de la década de 1970 un estudiante graduado del laboratorio de Hideki Shirakawa, del Instituto de Tecnología de Tokio, estaba intentando obtener poliacetileno a partir de acetileno ordinario (el gas utilizado en las soldaduras). El polímero, un polvo oscuro, se había sintetizado en 1955, pero nadie sabía demasiado sobre él. En lugar de un polvo oscuro, el estudiante de Shirakawa obtuvo una lámina brillante y plateada de aspecto parecido al aluminio, fácilmente extensible. Al repasar sus fórmulas, el estudiante encontró un error; la cantidad de catalizador utilizado era 1000 veces mayor que la indicada en las instrucciones. Había producido efectivamente poliacetileno, pero en una forma diferente de la de cualquier poliacetileno fabricado hasta entonces.

Cuando uno de los autores de este artículo (MacDiarmid) visitó el laboratorio de Shirakawa en 1976, acababa de comenzar la búsqueda de los metales sintéticos, en cuyo contexto el poliacetileno constituía un claro objetivo de investigación. Shirakawa pasó luego un año en la Universidad de Pennsylvania con MacDiarmid y su colega Alan J.Heeger, estudiando las posibilidades del nuevo polímero. La colaboración dio sus frutos cuando los tres investigadores intentaron dopar el material con iodo. Las láminas plateadas flexibles se convirtieron en láminas metálicas doradas; ¡la conductividad del poliacetileno aumentó en más de mil millones!

Desde aquellas fechas se ha visto que una docena de polímeros y derivados experimentan esta transición fortuita cuando se dopan. Todos ellos están constituidos fundamentalmente por átomos de carbono y átomos de hidrógeno dispuestos en "unidades monoméricas repetidas", unidas entre sí para formar los polímeros [*véase la figura 2*]. Algunas de estas unidades repetidas poseen átomos de nitrógeno o azufre, además del carbono e hidrógeno. El más sencillo de los polímeros conductores, el poliacetileno, está constituido por una cadena continua de carbono en la que a cada átomo de carbono se le une un átomo de hidrógeno. Los átomos de carbono están unidos entre sí por una serie alternativa de enlaces simples y dobles que intercambian sus posiciones continuamente a lo largo de la cadena de carbono, dando de hecho lugar a un enlace intermedio entre los enlaces simple y doble [*véase la figura 3*]. Esta configuración conjugada resulta característica de los polímeros conductores.

Aunque presentan algunas semejanzas estructurales, las propiedades de los polímeros pueden ser muy diferentes entre sí. Su sensibilidad al calor y al aire, su solubilidad y la facilidad con



1. CAMBIOS DE COLOR que acompañan la síntesis del poliacetileno, un polímero que se hace conductor al ser dopado o contaminado. La película delgada es de color rojo a la luz transmitida cuando se polimeriza sobre las paredes de una vasija de reacción; se vuelve azul al calentar dicha vasija. El hielo seco (*abajo*) y los elementos de calefacción (*arriba*) mantienen el poliacetileno a dos temperaturas diferentes para poner de manifiesto el cambio de color. Las láminas más gruesas tales como la representada en la figura 5 reflejan la luz, en lugar de transmitirla, y brillan como los metales. (Ilustración de James Kilkelly.)

que se pueden moldear, así como su conductividad, varían bastante según el tipo de polímero y el método utilizado para sintetizarlo. Por su especial combinación de características, el poliparafenileno, el politiofeno, el polipirol y la polianilina han recibido una gran atención, pero el poliacetileno se considera todavía el prototipo de los polímeros conductores y es el más profundamente estudiado. La conductividad de este material es más alta que la de cualquier otro polímero: el pasado año, los investigadores de la empresa B.A.S.F. A.G. de la República Federal Alemana encontraron que al contaminar una forma muy pura de poliacetileno habían obtenido un material cuya conductividad era la cuarta parte de la conductividad del cobre a igualdad de volumen y el doble que la de éste a igualdad de peso. ¿Qué mecanismos moleculares explican este comportamiento?

Los cambios electrónicos y físicos precisos que se producen en un proceso de dopado son todavía objeto de debate, pero se conocen ya algunos aspectos fundamentales del proceso. La conductividad se mide en mhos por centímetro, es decir, el inverso de la unidad de resistividad, ohms por centímetro. La diferencia entre el estado aislante y el conductor es enorme: la conductividad de los buenos aislantes, el teflon y el poliestireno por ejemplo, son del orden de 10^{-18} mhos por centímetro; los buenos conductores como el cobre y la plata tienen conductividades del orden de 10^6 mhos por centímetro. El poliacetileno de la B.A.S.F. tiene una conductividad de unos 147.000 mhos por centímetro.

La electricidad circula gracias al movimiento de los electrones; en consecuencia, para que un material conduzca la electricidad es necesario que algunos de sus electrones se muevan libremente a través del mismo. En los sólidos que poseen una extensa red de enlaces atómicos, los electrones "se mueven" en el interior de ciertos estados discretos de energía conocidos como bandas. Cada banda de energía posee una capacidad definida para albergar electrones, bandas que pueden estar también vacías. Los electrones han de poseer cierta energía para ocupar una banda determinada. Para que un electrón pueda saltar de una banda a otra de energía superior, necesita recibir energía.

El movimiento de los electrones exige también la existencia de una banda parcialmente llena. Las bandas vacías no pueden conducir la electricidad y lo mismo sucede con las llenas. La conducción se parece a una carrera de relevos en la que los electrones se mueven a través de las bandas de energía como si fueran calles. Para que se produzca el paso de la corriente, es necesario que otro "corredor" esté esperando en la pista; por tanto, la banda no puede quedar vacía. Los electrones deben poder entrar en la banda donde los otros corredores esperan; por consiguiente, la banda en cuestión tampoco puede estar llena.

Los metales son buenos conductores de la electricidad porque poseen bandas de energía parcialmente llenas. Los aislantes y los semiconductores son materiales cuyas bandas de energía están completamente llenas o completamente vacías. La banda ocupada más elevada se denomina banda de valencia y, banda de conducción, la banda vacía situada inmediatamente por encima de la anterior. El intervalo de energía que separa estas dos bandas en los aislantes es muy grande; en el caso de los semiconductores, dicho intervalo es algo menor. Tan sólo los electrones que reciben impulsos de energía procedentes de fuentes de calor o de luz pueden cruzar dicho intervalo.

El perfil electrónico de los polímeros ordinarios es parecido al de los aislantes y semiconductores: su banda de valencia está llena, su banda de conducción se halla vacía y las dos bandas quedan separadas entre sí por cierto intervalo energético. El dopado altera este perfil tomando electrones de la banda de valencia (en el caso del dopado-*p*, o positivo, dado que la molécula adquiere una carga positiva) o cediendo electrones a la capa de conducción (en el caso del dopado-*n*, o negativo, dado que la molécula adquiere una carga negativa). En consecuencia, el polímero se convierte en conductor porque la banda de conducción o la banda de valencia se llenan parcialmente.

En realidad, el proceso de transición es algo más complicado. La carga introducida en la cadena de un polímero mediante dopado produce un cambio pequeño, aunque importante, en la po-

CARBONO
HIDROGENO
S AZUFRE
N NITROGENO

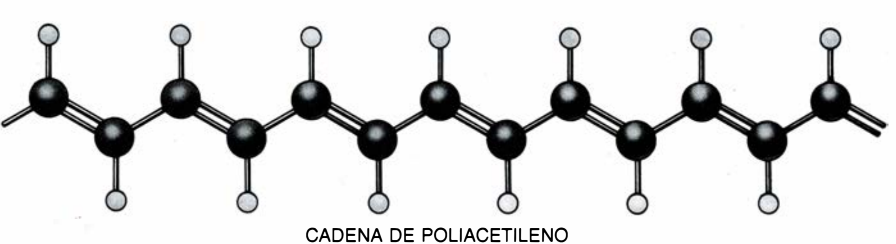
POLIACETILENO

POLIPARAFENILENO

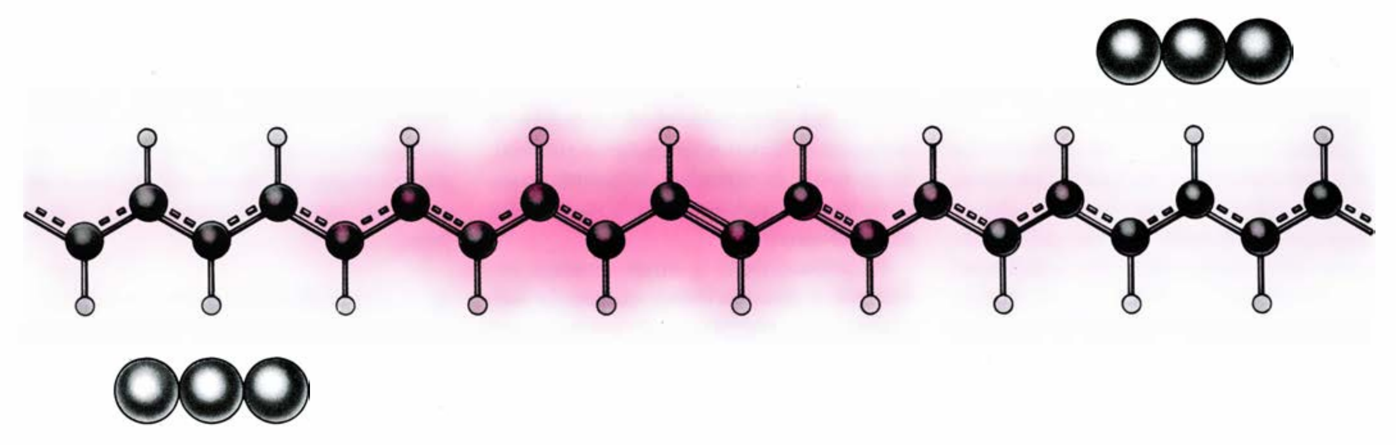
POLIPIRROL

POLITIOFENO

POLIANILINA


CADENA DE POLIACETILENO

2. UNIDADES MONOMERICAS REPETIDAS: se unen entre sí y forman las cadenas que dan lugar a los polímeros que se indican debajo de cada unidad. La característica más importante de estas moléculas es la existencia de enlaces simples y dobles alternados, lo cual crea una configuración susceptible de ser dopada. En la parte superior, se representan las unidades básicas de los cinco polímeros más populares; en la inferior, una cadena de poliacetileno sin dopar. En su forma no contaminada, estos polímeros no son conductores.



3. CADENA DE POLIACETILENO *P*-DOPADO: tiene una carga positiva deslocalizada que le permite conducir la electricidad como si fuera un metal. La densidad de electrones (*color*) se distribuye a lo largo de la cadena en forma de nube internamente estructurada, cuya densidad disminuye en las proximidades del ion contaminante negativo. La diferencia entre los enlaces doble y simple se diluye cuando nos acercamos al contaminante (*indicado por las líneas a trazos*).

sición de sus átomos. Este cambio favorece la formación de una especie de "islas" de carga conocidas por solitones, polarones y bipolarones. Estas islas se forman cerca de los iones contaminantes para niveles de dopado elevados; cuando existen unos 15 átomos de carbono por cada molécula de dopado, las islas empiezan a solaparse unas con otras. Las islas contiguas crean nuevas bandas de energía entre las bandas de valencia y de conducción, haciendo que los electrones puedan moverse con entera libertad entre ellas. Los solitones, polarones y bipolarones son los responsables de la conductividad de los polímeros.

Las bandas de energía singulares de los polímeros conductores no existen en los semiconductores dopados. Sin embargo, estas bandas pueden desempeñar cierto papel en el mecanismo de conducción de los modernos superconductores a temperaturas elevadas. Por otro lado, el descubrimiento de los solitones en el poliacetileno cargado ha despertado el interés de los físicos teóricos de todo el mundo, debido a que el polímero conductor proporciona un sistema excelente para estudiar este fenómeno único.

La relación entre la molécula dopante y el material contaminado difiere también en el caso de los semiconductores clásicos y los polímeros. Citemos un ejemplo: el dopado del silicio consiste en la sustitución de algunos de los átomos de silicio por átomos que posean un número mayor o menor de electrones que este elemento. En el dopado de polímeros conductores, las moléculas contaminantes nunca reemplazan a los átomos del polímero, sino que actúan como meros agentes receptores o dadores de electrones a las bandas de energía del polímero. Para niveles de contaminación elevados, cuando se forman las islas de solitones, polarones o bipolarones, la carga creada sobre el polímero está más deslocalizada y se comparte entre muchos átomos de carbono.

Dado que todavía subsisten muchas incógnitas acerca de los detalles del transporte de electrones a lo largo de las cadenas, los investigadores son incapaces de explicar el paso de los electrones de una cadena a otra. Sin embargo, parece indudable que ese tránsito se produce, puesto que una cadena determinada no recorre toda la longitud de una película de polímero. Es posible que exista algún tipo de conexión entre las cadenas; de manera alternativa, en el caso del poliacetileno dopado a niveles medios, la carga puede transportarse por medio de un móvil especial, una red de solitones "líquidos" que empapa las cadenas de polímeros más próximas.

Ninguna teoría puede explicar por qué razón los polímeros conductores poseen una gama tan amplia de conductividades. Sin embargo, los investigadores se han familiarizado empíricamente con los polímeros hasta tal punto que han llegado a descubrir los factores que afectan a la conductividad. El más evidente es el porcentaje de contaminante; este factor puede controlarse con precisión. Otro factor es la orientación de las cadenas de polímero en el material. La estructura de las cadenas da lugar a cierta anisotropía o asimetría: fenómeno en el que una cadena conduce mejor la electricidad en una dirección que en otra. Por ejemplo, en el poliacetileno dopado la conductividad puede ser hasta 1000 veces mayor a lo largo de la cadena que en una dirección perpendicular a ella.

La conductividad de las películas de polímero contaminado puede aumentarse alineando las cadenas de forma paralela. En realidad, esta alineación se realiza parcialmente en el poliacetileno de la B.A.S.F. Una vez sintetizado el polímero, y antes de doparse, se procede a estirar las láminas de poliacetileno, con el fin de orientar las fibras [*véase la figura 4*].

La pureza afecta también a la conductividad. Las impurezas pueden entorpecer la movilidad de los electrones, produciendo defectos que impiden el paso de electrones a través de las cadenas del polímero. Muchos investigadores atribuyen las propiedades del polímero de la B.A.S.F. al proceso de fabricación, que proporciona un polímero de elevada homogeneidad y pureza.

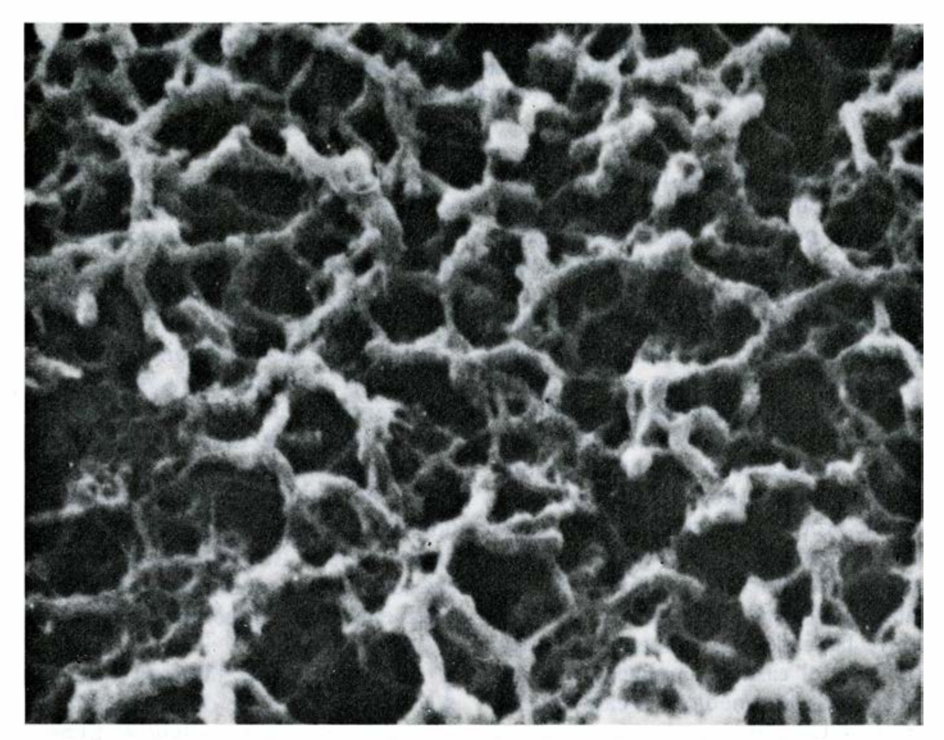
Los polímeros se fabrican y contaminan mediante diferentes procesos químicos o electroquímicos. El proceso estándar de producción de poliacetileno es de tipo químico y no difiere demasiado del utilizado por el estudiante de Shirakawa. El gas acetileno se hace llegar a un recipiente de vidrio recubierto con un catalizador químico que favorece la polimerización. Sobre el vidrio se inicia la formación de una película coherente; al cabo de cinco minutos, se ha formado ya una capa de poliacetileno no contaminado tan gruesa como una hoja de papel. Después de lavarse, la película se desprende de las paredes de la vasija; se provoca a continuación un dopado de tipo *p* mediante un tratamiento con una disolución de iodo, o una contaminación de tipo *n*, mediante una disolución de sodio metálico en mercurio.

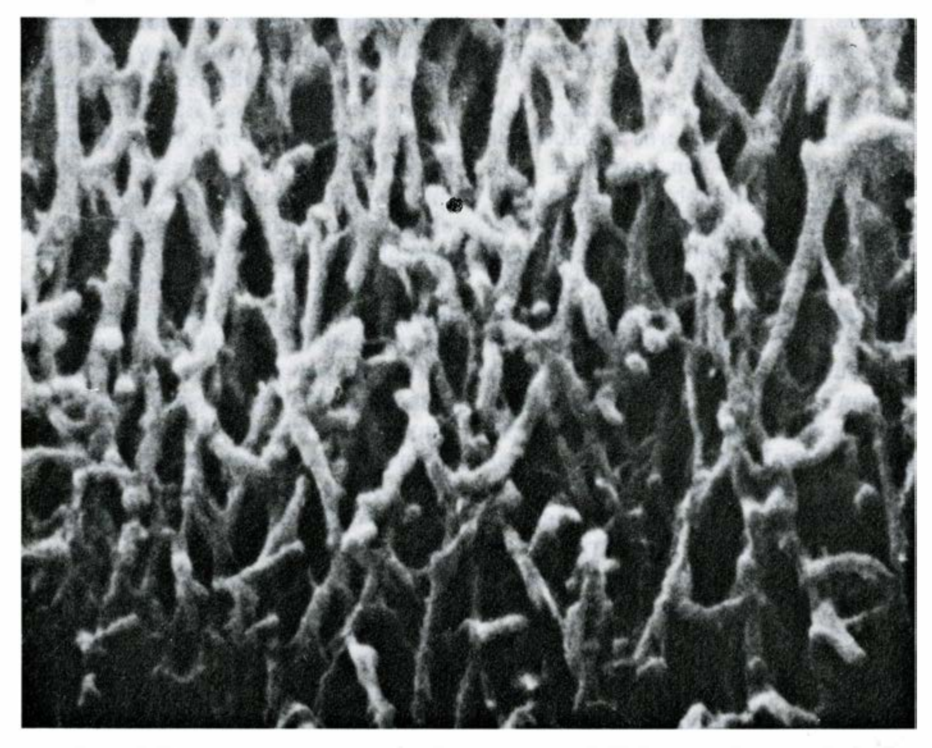
Con las técnicas electroquímicas, la síntesis y el dopado se producen simultáneamente. Dos electrodos metálicos están sumergidos en una disolución que contiene el ion contaminante y el monómero que constituirá el polímero. En un síntesis de dopado *p* típico, se extraen electrones de los monómeros adyacentes al electrodo positivo, gracias

a la aplicación de un potencial eléctrico adecuado. Este hecho da lugar a que los monómeros se polimericen sobre la superficie del electrodo.

Durante el proceso de polimerización, los electrones se siguen extrayendo del polímero recién formado, que se carga positivamente y atrae, por tanto, a los iones negativos —los iones contaminantes— presentes en la disolución. Sobre el electrodo positivo se forma una película de polímero dopado; ésta puede luego desprenderse. El porcentaje de contaminante depende de la cuantía de la carga positiva que transporte el polímero, que, a su vez, depende del potencial final aplicado al polímero. Ajustando dicho potencial, se controla la cantidad de contaminante presente en el polímero y, por tanto, su conductividad.

El poliacetileno y otros muchos polímeros se pueden dopar utilizando métodos electroquímicos, después de sintetizarse por métodos químicos, uniendo unas cintas del polímero a un electrodo positivo o negativo sumergido en una disolución que contenga el ion dopante. Mejorando la estructura, la pureza y los métodos de contaminación, la conductividad de los polímeros superará algún día la del cobre, tanto en relación con el volumen, como con el peso.





4. AL ESTIRAR LAS PELICULAS DE POLIMEROS, las fibras se alinean y la conductividad aumenta en la misma dirección del estiramiento. Estas microfotografías electrónicas muestran la estructura del poliacetileno antes (*arriba*) y después (*abajo*) del proceso. Cada fibra consta de unas 1500 cadenas de polímero.

La conductividad constituye sólo uno de los criterios que determinan si un material resulta interesante en la práctica. Cuando se logró la síntesis de los polímeros conductores, se pensó que algunas de sus propiedades físicas y químicas podrían limitar sus aplicaciones tecnológicas. Los polímeros conductores eran insolubles y no se podían fundir, ni moldearse, por tanto, como los polímeros ordinarios. Por otra parte, muchos de ellos se descomponían al exponerse al aire, hecho que limitaba considerablemente sus aplicaciones potenciales.

En estos dos últimos años el primero de estos problemas se ha resuelto incorporando en la cadena de carbono otros grupos químicos capaces de modificar las propiedades de la molécula. Por ejemplo, el trabajo llevado a cabo en la empresa Allied-Signal ha demostrado que el politiofeno se puede volver soluble en disolventes comunes reemplazando un átomo de hidrógeno por grupos de hidrocarburos grandes, como el carbono—cuatro butilo. Por otra parte, cuando una cadena de 12 carbonos se une a cada una de las unidades monoméricas repetidas, el polímero se puede fundir y preparar en forma de fibras que, al doparse con iodo, resultan buenos conductores de la electricidad.

Gracias a estos avances, se han multiplicado los posibles usos de los polímeros conductores. La aplicación más imminente y conocida es la batería recargable. A principios de nuestra década, MacDiarmid y Heeger demostraron que el poliacetileno con dopado *n* y *p* podía servir para fabricar electrodos de un prototipo de batería recargable. Los polímeros fueron presentados como el invento que permitiría convertir en realidad los coches eléctricos, dado que el peso de las baterías de polímeros sería mucho menor que el de las baterías convencionales de plomo y ácido.

Aunque los esfuerzos dedicados a desarrollar los coches eléctricos han disminuido, no ha decaído el interés por las baterías recargables con electrodos de polímero. Se cree que los electrodos de polímero durarían mucho más que los metálicos, debido a que los iones implicados en el suministro y al-

macenamiento de la carga eléctrica provienen de la disolución, y no de los propios electrodos. En consecuencia, los electrodos de polímero evitarían el desgaste mecánico asociado con la disolución y deposición del material de los electrodos que tienen lugar durante los ciclos de carga y descarga de las baterías ordinarias. Por otra parte, al contrario de lo que sucede con las baterías de níquel-cadmio y de plomo-ácido, las baterías de polímeros no contienen sustancias tóxicas. Por consiguiente, los problemas relacionados con su manejo serían mínimos.

Las empresas Bridgestone y Seiko han desarrollado conjuntamente una batería recargable de tipo moneda con un electrodo de polianilina que se puso a la venta en Japón meses atrás. El otro electrodo está fabricado con litio metálico. Se dice que tales baterías triplican la capacidad de las baterías recargables de litio existentes en la actualidad, con un voltaje de dos a tres veces mayor que el de las baterías de níquel-cadmio y 1,5 veces mayor que el de las baterías de plomo-ácido que se utilizan en los automóviles. La velocidad de descarga espontánea, que determina la vida de una batería durante su almacenaje, resulta ser significativamente menor que la de las baterías de níquel-cadmio y plomo-ácido.

En la República Federal Alemana, la B.A.S.F. y la empresa Varta están experimentando con una batería recargable flexible con un electrodo de litio metálico y el otro de polipirol. Uno de los modelos experimentales es una célula del tamaño de una tarjeta postal, de un grosor de unos cuatro milímetros, que suministra tres volt; se dobla como si fuera una cartulina. Allied-Signal está trabajando también en una batería recargable con un electrodo de polímero conductor. Esta compañía no ha hecho públicos los pormenores de su funcionamiento, pero parece ser que la batería en cuestión almacena casi el doble de energía eléctrica que una batería de niquel-cadmio similar, a igualdad de peso.

El progreso alcanzado con los electrodos de polímeros es bastante alentador, pero las posibilidades de los polímeros conductores van mucho más allá de las baterías. Allied-Signal ya tiene algunos proyectos en marcha para explotar la aplicación de los polímeros conductores en dispositivos sensores. En este campo, se podrían seguir diferentes estrategias. Algunos polímeros dopados se descomponen con el calor, sufriendo una pérdida de conductividad. Conectados con un medidor de resistencias, estos polímeros permitirían controlar la exposición de productos farmacéuticos a temperaturas que pudieran alterar su eficacia. Los polímeros conductores se han incorporado ya en ciertos sensores utilizados para poner de manifiesto si los alimentos congelados se han descongelado y vuelto a congelar durante su manipulación. De modo análogo, es posible detectar la presencia de sustancias tóxicas que actúan como dopantes de polímeros no contaminados. Los polímeros podrían actuar también como sensores de radiación si se colocaran en una atmósfera de gases que se convirtieran en dopantes activos al ser expuestos a la radiación.

Varios países están investigando la utilización de politiofeno y polianilina en pantallas electrocrómicas que cambian de color cuando se les aplica un potencial eléctrico. Por ejemplo, las láminas delgadas de politiofeno son rojas en el estado dopado y cambian a azul intenso en el estado no contaminado. Entre las aplicaciones podrían citarse las pantallas de horarios en aeropuertos y estaciones, carteles publicitarios en el interior de los edificios, así como las calculadoras, los ordenadores, los relojes y cualquier otro dispositivo de los que utilizan actualmente pantallas de cristal líquido (LCD). Aunque estos materiales podrían tener una amplia gama de colores, es difícil decir ahora si los dispositivos electrocrómicos resultarían más ventajosos que los cristales líquidos. Para que estos dispositivos fueran prácticos, debería mejorarse sustancialmente el tiempo de conexión-desconexión y la vida operativa de los polímeros.

El mismo mecanismo que hace que los polímeros conductores cambien de color en un dispositivo electrocrómico podría utilizarse también en los cristales de las ventanas. Unas capas muy delgadas de polímeros, incrustadas en un electrolito sólido incoloro y colocadas entre dos láminas de vidrio, podrían oscurecer una ventana como resultado de la aplicación de un potencial eléctrico. El grado de oscurecimiento se podría regular en función del valor del potencial aplicado. Por otra parte, el hecho de que los polímeros conductores sean capaces de absorber la luz solar ha atraído la atención de los investigadores que trabajan en células de energía solar. Sin embargo, el rendimiento de conversión alcanzado hasta ahora resulta desalentador.

Los polímeros conductores absorben también energía electromagnética a bajas frecuencias y podrían utilizarse, por



5. LAMINA DE POLIACETILENO desprendida de las paredes de una vasija de reacción al concluir el proceso de síntesis. Esta lámina, flexible como un papel, debe doparse para convertirla en conductora.

tanto, como escudos electromagnéticos para detener las pérdidas de radiación en los terminales de ordenadores. Con este fin, se utilizan normalmente plásticos rellenos de carbono o metales, pero la facilidad de manipulación y la mayor conductividad de los polímeros dopados homogéneamente les proporcionan unas especiales ventajas. Las elevadas pérdidas de microondas encontradas recientemente en la polianilina resultan prometedoras para esta aplicación particular.

Los polímeros dopados se han utilizado también como líneas de conducción en los modernos circuitos impresos, en transistores y como sustitutos de las uniones de semiconductores convencionales; por ahora, su respuesta resulta ser inferior a la de los materiales convencionales. Sin embargo, los avances experimentados en la síntesis de polímeros, con el fin de obtener moléculas mejor alineadas y con menos defectos, así como los progresos en las técnicas de dopado, proporcionarán mejoras sustanciales en su funcionamiento. A consecuencia de sus peculiares propiedades ópticas, se está considerando la aplicación de los polímeros conductores en la próxima generación de ordenadores ópticos.

El cuerpo humano es otro "dispositivo" en el que los polímeros conductores podrían desempeñar un papel en el futuro. Dado su carácter inerte y estable, se especula con la utilización de ciertos polímeros para prótesis neurológicas (nervios artificiales). En particular, el polipirol podría resultar adecuado por su carácter no tóxico y su capacidad de suministrar una carga eléctrica adecuada. El ion contaminante podría ser, en este caso, la heparina, un producto químico que impide la coagulación de la sangre y que actúa de forma bastante adecuada como dopante del polipirol. Por otro lado, los polímeros podrían actuar como sistemas internos de suministro de fármacos, implantándolos en el interior del cuerpo y contaminándolos con moléculas de dichos productos. Los fármacos en cuestión se liberarían cuando el polímero se transformara en su estado neutro mediante la aplicación debidamente programada de un potencial eléctrico.

En muchos aspectos, la situación de los polímeros conductores a mediados de los años 80 es similar a la de los polímeros convencionales hace medio siglo. Aunque los polímeros convencionales se sintetizaron y estudiaron en muchos laboratorios de todo el mundo, no se convirtieron en materiales de utilidad técnica hasta que sufrieron ciertas modificaciones químicas cuyo desarrollo duró varios años. De forma análoga, las propiedades químicas y físicas de los polímeros conductores deben ajustarse a cada aplicación si se pretende que los productos obtenidos sean económicamente rentables. Al margen de las posibles aplicaciones que se puedan descubrir en el futuro para los polímeros conductores, estos materiales constituirán un reto para la investigación básica, a medida que se vayan descubriendo nuevos e inesperados fenómenos. Sólo el tiempo nos dirá si el impacto de estos plásticos conductores igualará el de sus parientes aislantes.

Ramanujan y el número pi

Hace 75 años, un genio matemático indio ideó métodos de extraordinaria eficacia para calcular el valor de pi. En la actualidad, su procedimiento forma parte de algoritmos que rinden millones de cifras decimales de pi

Jonathan M. Borwein y Peter B. Borwein

El número pi, que es la razón de la circunferencia de un círculo cualquiera al diámetro del mismo, se calculó en 1987 con una precisión sin precedentes: más de 100 millones de cifras decimales. El año pasado se cumplió también el centenario del nacimiento de Srinivasa Ramanujan, genio matemático indio, bastante enigmático, que pasó gran parte de su breve vida solo y enfermo. La verdad es que ambos acontecimientos están estrechamente emparentados, porque el método básico que subyace a los cálculos más recientes de pi lo ideó Ramanujan, por mucho que su puesta en práctica hubiera de esperar a la formulación de los correspondientes algoritmos (lo que han conseguido diversos investigadores, entre ellos, los autores), al advenimiento de los modernos superordenadores y a la invención de nuevos procedimientos para la multiplicación de números.

Aparte de constituir un campo donde establecer marcas exóticas, el empeño puesto en determinar millones de cifras decimales del número pi parece, a primera vista, bastante fútil. Bastarían 39 cifras decimales de pi para calcular con error menor que el radio de un átomo de hidrógeno el perímetro de una circunferencia capaz de abarcar la totalidad del universo conocido. Cuesta imaginar situaciones físicas que requieran mayor número de cifras decimales. ¿Por qué razón no se dan por satisfechos los matemáticos con los 50 primeros decimales de pi, por poner una cifra?

Son varias las respuestas que pueden darse. Una, que el cálculo de pi se ha convertido en una especie de banco de pruebas computacional: proporciona una medida del refinamiento y fiabilidad de los ordenadores que lo efectúan. Por otra parte, la búsqueda de valores de pi cada vez más precisos lleva a los matemáticos a desconcertantes e inesperados reductos de la teoría de números. Otro motivo, más ingenuo, es sencillamente que el problema "está ahí". Y en efecto, desde hace más de dos milenios y medio, el número pi viene constituyendo un elemento permanente de la cultura matemática.

Además, siempre cabe la posibilidad de que tales cálculos arrojen luz sobre algunos de los misterios que rodean a pi, una constante universal que, a pesar de lo relativamente elemental de su naturaleza, no acaba de comprenderse. Por ejemplo, aunque se ha demostrado que jamás podrá evaluarse pi con exactitud sometiendo enteros positivos a una combinación de operaciones de adición, sustracción, multiplicación, división y extracción de raíces, hasta la fecha no se ha probado que las cifras de pi sigan una distribución aleatoria (y, por tanto, que todas las cifras, de 0 a 9, aparezcan con la misma frecuencia). Cabe en lo posible, aunque es sumamente improbable que, a partir de un momento dado, todos los dígitos de pi sean exclusivamente 0 o 1, o que presenten alguna otra regularidad. Además, pi aparece en toda clase de lugares inesperados, que nada tienen que ver con las circunferencias. Por ejemplo, tomando un número al azar en el conjunto de todos los enteros, la probabilidad de que tal número carezca de factores primos repetidos es seis dividido por el cuadrado de pi. Igual que otros eminentes matemáticos, Ramanujan quedó cautivado por la fascinación que ejerce el número pi.

Los ingredientes de las modernas técnicas de cálculo de pi se cuentan entre los tesoros matemáticos que han ido aflorando a causa del renovado interés por la obra de Ramanujan. No obstante, gran parte de su obra sigue inaccesible a los investigadores. El cuerpo principal de los trabajos de Ramanujan está recogido en sus "Cuadernos", conjuntos de anotaciones personales que Ramanujan redactó con una nomenclatura propia y particular. Por si los matemáticos que han estudiado los "Cuadernos" no tuvieran suficientes dificultades, se da la circunstancia de que Ramanujan no solía consignar demostraciones formales de sus teoremas. Parece próxima a concluir la tarea de desciframiento y preparación de esos "Cuadernos" que está realizando Bruce C. Berndt, de la Universidad de Illinois en Urbana-Champaign.

Que sepamos, jamás se ha realizado una recopilación matemática de alcance y dificultad semejantes. El empeño ciertamente valdrá la pena. La herencia que Ramanujan dejó en sus "Cuadernos" no sólo promete enriquecer a la matemática pura, sino que hallará aplicaciones en diversos campos de la física matemática. Por ejemplo, Rodney J. Baxter, de la Universidad Nacional Australiana, reconoce que los descubrimientos de Ramanujan le ayudaron a resolver problemas de mecánica estadística, como el llamado modelo del hexágono duro, que estudia el comportamiento de un sistema de partículas que interactúan entre sí y se hallan repartidas sobre una rejilla similar a un panal. Análogamente, Carlos J. Moreno, de la Universidad de la Ciudad de Nueva York, y Freeman J. Dyson, del Instituto de Estudios Avanzados, han señalado que los físicos están comenzando a aplicar trabajos de Ramanujan en la teoría de supercuerdas.

La talla matemática de Ramanujan nos deja todavía más asombrados habida cuenta de lo reducida que fue su educación formal. Nació el 22 de diciembre de 1887 en Erode (India meridional) en el seno de una familia venida a menos de la casta de los brahmanes. Se crió en Kumbakonam, donde su padre ejercía de contable de un pañero. Pronto se reconoció su precocidad matemática, concediéndosele, a la edad de siete años, una beca para asistir a la escuela pública del lugar. Según parece, les recitaba fórmulas matemáticas a sus compañeros de clase, entre otras, muchas cifras decimales del número pi.

A los doce años, Ramanujan dominaba el contenido del tratado de trigonometría de S. L. Loney, *Plane Trigonometry*, obra bastante completa sobre el tema, en la que figuraba un análisis de las sumas y productos de sucesiones infinitas que más adelante desempeñarían en sus trabajos tan prominente papel. (Una sucesión infinita es una ristra de términos que no tiene fin, a menudo generados mediante una sencilla fórmula. En el contexto que nos ocupa, las sucesiones de interés son aquellas cuyos términos producen, al sumarlos o multiplicarlos, un valor finito y perfectamente identificable. En el caso de que se proceda a sumar los sucesivos términos de una sucesión, la expresión resultante se denomina serie; y si a multiplicarlos, producto.) Tres años después le prestaron la *Synopsis of Elementary Results in Pure Mathematics*, una relación de unos 6000 teoremas (la mayoría de los cuales venían sin demostración) recopilada por G. S. Carr, un profesor de la Universidad de Cambridge. Esos dos libros constituyeron la formación matemática básica de Ramanujan.

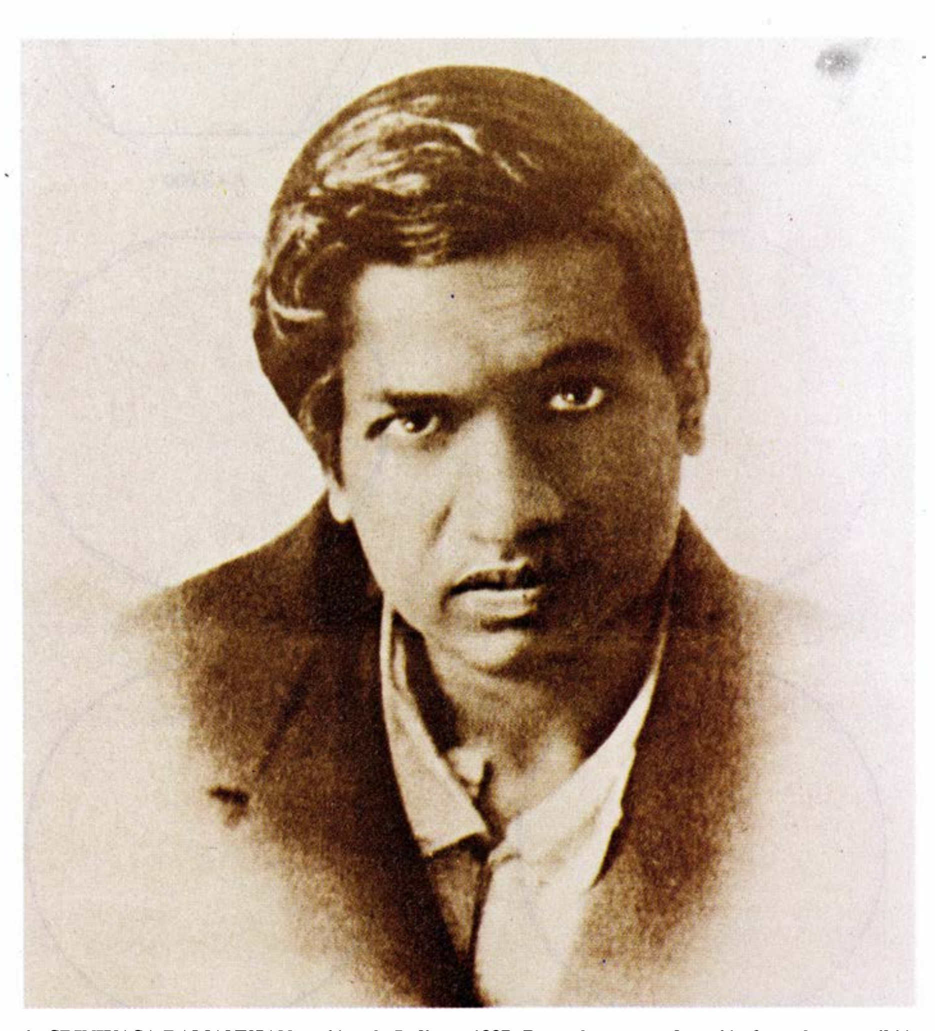
En 1903 Ramanujan fue admitido en un colegio universitario de la localidad. Sin embargo, la absoluta dedicación a sus diversiones matemáticas le hizo fracasar en los exámenes, situación que se repitió cuatro años después en otro centro universitario de Madrás. Ramanujan sí dejó a un lado su vocación —al menos temporalmente— para buscar un empleo tras contraer matrimonio, en 1909. R. Ramachandra Rao, un adinerado mecenas matemático, le concedió un estipendio mensual basándose en las favorables recomendaciones de varios matemáticos indios y en los hallazgos que Ramanujan había anotado ya en sus "Cuadernos".

En 1912, deseando un trabajo más convencional, ocupó una plaza en las oficinas de la Junta del Puerto de Madrás, presidida por un ingeniero británico, Sir Francis Spring, y cuyo director gerente era V. Ramaswami Aiyar, fundador de la Sociedad Matemática India. Ambos animaron a Ramanujan para que comunicase sus resultados a tres distinguidos matemáticos británicos. Según parece, dos de ellos no le respondieron; sí le contestó G. H. Hardy, de Cambridge, tenido en nuestros días por el más eminente de los matemáticos británicos de la época.

Hardy, habituado como estaba a recibir cartas de excéntricos convencidos de ser grandes matemáticos, estuvo a punto de desechar a primera vista la carta de Ramanujan el día mismo que la recibió, el 16 de enero de 1913. Pero aquella noche, después de la cena, Hardy y un colega amigo suyo, John E. Littlewood, se sentaron a descifrar una lista de 120 fórmulas y teoremas que Ramanujan había añadido a la carta. Algunas horas más tarde habían llegado a un veredicto: tenían ante sí la obra de un genio, no la de un chiflado. (Según su propia "escala de talento puro" con la que graduaba a los matemáticos, más tarde Hardy concedería un 100 a Ramanujan, un 30 a Littlewood y un 25 a sí mismo. El matemático alemán David Hilbert, el más influyente de su época, sólo merecería un 80.) Hardy describió la revelación y sus consecuencias como el incidente más romántico de su vida. Escribió que algunas de las fórmulas de Ramanujan le desbordaron por completo y que, ello no obstante, "forzoso era que fueran verdaderas, porque de no serlo, nadie habría tenido la imaginación necesaria para inventarlas".

Hardy invitó inmediatamente a Ramanujan a Cambridge. Pese a las fuertes objeciones de su madre, por no hablar de sus propias reservas, Ramanujan partió para Inglaterra en marzo de 1914. Durante los cinco años siguientes, Hardy y Ramanujan trabajaron codo con codo en el Trinity College. La destreza técnica de Hardy, unida a la brillantez "en rama" de Ramanujan, fructificaron en una colaboración sin par. Publicaron conjuntamente una serie de artículos seminales sobre las propiedades de diversas funciones aritméticas y prepararon el terreno para afrontar problemas como: ¿cuántos divisores primos es probable que tenga un número dado? ¿De cuántas maneras distintas puede expresarse un número en forma de suma de enteros positivos menores que él?

En 1917 Ramanujan fue admitido como miembro numerario de la Royal Society de Londres y también del Trinity College, siendo el primer indio al



1. SRINIVASA RAMANUJAN nació en la India en 1887. Pese a la escasa educación formal que recibió, reconstruyó casi por sí solo gran parte del edificio de la teoría de números y lo llevó a nuevas alturas, aportándole numerosas fórmulas y teoremas originales. Como tantos otros ilustres matemáticos antes que él, Ramanujan sucumbió a la fascinación del número pi, que es la razón entre la circunferencia de un círculo cualquiera y su diámetro. Basándose en su investigación de las funciones modulares [*véase el recuadro de la página 76*], Ramanujan dio expresiones exactas de pi y dedujo para ellas valores aproximados. Gracias a los trabajos de diversos investigadores (entre ellos, los de los autores) se tiene hoy una comprensión más completa de los métodos de Ramanujan, que han sido llevados a la práctica integrados en algoritmos de cálculo.

que se concedía tal honor. Empero, conforme crecía su importancia, su salud se deterioraba gravemente, en un declive acelerado tal vez por la dificultad de mantener una dieta estrictamente vegetariana en una Inglaterra sometida a racionamiento a causa de la guerra. Sus reiteradas visitas a sanatorios no impidieron que Ramanujan mantuviera la producción de nuevos resultados. En 1919, cuando la paz devolvió la seguridad a los viajes, Ramanujan regresó a la India. Convertido en ídolo de los jóvenes intelectuales indios, Ramanujan murió el 26 de abril de 1920, de una enfermedad diagnosticada a la sazón como tuberculosis, pero que, según parece hoy, debió estar causada por una grave deficiencia vitamínica. Fiel hasta el fin a las matemáticas, Ramanujan no aflojó el paso a pesar de los sufrimientos de los últimos meses de su vida, y produjo una notable obra recogida en el llamado "Cuaderno perdido".

El trabajo de Ramanujan sobre el número pi fue fruto, en buena medida, de sus investigaciones sobre las ecuaciones modulares, posiblemente el tema que más a conciencia abordan sus "Cuadernos". A grandes rasgos, una ecuación modular es una relación algebraica entre una función expresada mediante una variable x —en notación matemática $f(x)$— y esa misma función dada a partir de potencias enteras de x; por ejemplo, $f(x^2)$, $f(x^3)$ o $f(x^4)$. El "orden" de la ecuación modular está determinado por el exponente de la potencia entera. La ecuación modular más sencilla es la de segundo orden: $f(x) = 2\sqrt{f(x^2)}/[1 + f(x^2)]$. Evidentemente, no todas las funciones satisfarán una ecuación modular, pero existe una clase de funciones, las llamadas modulares, que sí la verifican. Estas funciones tienen diversas y sorprendentes propiedades de simetría que les reservan un lugar especial en las matemáticas.

Ramanujan no tuvo par en su capacidad para encontrar soluciones de ecuaciones modulares que verificasen además otras condiciones. Tales soluciones reciben el nombre de valores singulares. Resulta que, en ciertos casos, al resolver la ecuación en busca de valores singulares surgen números cuyos logaritmos naturales coinciden con pi (multiplicado por una constante) hasta un número sorprendente de cifras decimales [*véase el recuadro de la página 76*]. Aplicando con extraordinario virtuosismo esa técnica general, Ramanujan produjo multitud de notables series infinitas, amén de expresiones explícitas, que son aproximaciones de pi. Algunas de ellas se dan en el único artículo formal que Ramanujan dedicó al tema, *Modular Equations and Approximations to π*, publicado en 1914.

Los esfuerzos de Ramanujan por obtener aproximaciones de pi forman parte de una tradición venerable. Las primeras civilizaciones indoeuropeas ya tenían conciencia de que el área del círculo es proporcional al cuadrado de su radio, y de que su circunferencia es directamente proporcional al diámetro. Sin embargo, resulta mucho menos claro cuándo se comprendió por vez primera que la razón de la circunferencia de un círculo cualquiera a su diámetro y la razón del área de cualquier círculo al cuadrado de su radio eran la misma constante, simbolizada en nuestros días

2. METODO DE ARQUIMEDES para el cálculo del número pi; consistía en inscribir y circunscribir polígonos regulares (polígonos cuyos lados tienen todos la misma longitud) en un círculo de diámetro unidad. Los perímetros de los polígonos inscritos y circunscritos servían de cota inferior y superior para el valor de pi. Para calcular los perímetros de los polígonos pueden utilizarse hoy las funciones seno y tangente, como se muestra aquí, pero Arquímedes tuvo que desarrollar relaciones equivalentes basadas en construcciones geométricas. Valiéndose de polígonos de 96 lados determinó que pi es mayor que $3\frac{10}{71}$ y menor que $3\frac{1}{7}$.

por la letra griega π. (El símbolo, del que toma nombre la constante, se incorporó tardíamente a las matemáticas; lo introdujo en 1706 el escritor y matemático inglés William Jones y lo popularizó el matemático suizo Leonhard Euler en el siglo XVIII.)

PRODUCTO DE WALLIS (1665)

$$\frac{\pi}{2}=\frac{2\times 2}{1\times 3}\times\frac{4\times 4}{3\times 5}\times\frac{6\times 6}{5\times 7}\times\frac{8\times 8}{7\times 9}\times\cdots=\prod_{n=1}^{\infty}\frac{4n^2}{4n^2-1}$$

SERIE DE GREGORY (1671)

$$\frac{\pi}{4}=1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\cdots=\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}$$

FORMULA DE MACHIN (1706)

$$\frac{\pi}{4}=4\arctan(1/5)-\arctan(1/239),\qquad \text{siendo } \arctan X = X-\frac{X^3}{3}+\frac{X^5}{5}-\frac{X^7}{7}+\cdots=\sum_{n=0}^{\infty}(-1)^n\frac{X^{(2n+1)}}{2n+1}$$

RAMANUJAN (1914)

$$\frac{1}{\pi}=\frac{\sqrt{8}}{9{,}801}\sum_{n=0}^{\infty}\frac{(4n)![1{,}103+26{,}390n]}{(n!)^4\,396^{4n}},\quad \text{siendo } n! = n\times(n-1)\times(n-2)\times\cdots\times 1 \text{ y } 0!=1$$

BORWEIN Y BORWEIN (1987)

$$\frac{1}{\pi}=$$

$$12\sum_{n=0}^{\infty}\frac{(-1)^n(6n)![212{,}175{,}710{,}912\sqrt{61}+1{,}657{,}145{,}277{,}365+n(13{,}773{,}980{,}892{,}672\sqrt{61}+107{,}578{,}229{,}802{,}750)]}{(n)!^3(3n)![5{,}280(236{,}674+30{,}303\sqrt{61})]^{(3n+3/2)}}$$

3. POR SUMAS O PRODUCTOS de términos de sucesiones infinitas apropiadas se obtienen valores de pi (dividido por una constante) o de su recíproco. Las dos primeras sucesiones, descubiertas por los matemáticos John Wallis y James Gregory, se cuentan seguramente entre las más conocidas, pero resultan inútiles para realizar el cómputo. Cien años de cálculo ininterrumpido en un superordenador programado para sumar o multiplicar los términos de una cualquiera de esas sucesiones ni siquiera proporcionarían 100 cifras exactas de pi. La fórmula descubierta por John Machin facilitó la evaluación de la constante, pues el cálculo diferencial permite expresar el arco tangente de un número, *x*, mediante la suma de los términos de una serie, cuya convergencia hacia el verdadero valor del arco tangente es tanto más rápida cuanto menor es *x*. Los cálculos del valor de pi realizados desde principios del siglo XVIII hasta principios de la década de 1970 se fundaron en variantes de la fórmula de Machin. La suma de la serie de Ramanujan converge hacia el verdadero valor de 1/π mucho más rápidamente: cada uno de los sucesivos términos aporta alrededor de 8 dígitos exactos más. La última serie, formulada por los autores, añade unas 25 cifras por término; el primer término (para el que *n* es igual a cero) proporciona un número que coincide con pi hasta 24 cifras decimales.

Arquímedes de Siracusa, el mayor matemático de la antigüedad, estableció rigurosamente la equivalencia de ambas razones en su tratado *Medición de un círculo*. Calculó asimismo un valor de pi basándose en principios matemáticos y no en la medición directa de la circunferencia, área y diámetro del círculo. En efecto, Arquímedes inscribió y circunscribió a un círculo (cuyo diámetro se suponía igual a la unidad) polígonos regulares (polígonos cuyos lados y ángulos son todos iguales) y consideró que los respectivos perímetros de tales polígonos eran cotas inferiores y superiores de los posibles valores de la circunferencia del círculo, que es numéricamente igual a pi [*véase la figura 2*].

Tal método de aproximación del valor de pi no era nuevo. La idea de inscribir polígonos de número de lados progresivamente mayor había sido propuesta ya por Antífono, y un contemporáneo de éste, Brisón de Heraclea, había añadido al procedimiento los polígonos circunscritos. La novedad de Arquímedes radicaba en la correcta determinación del efecto de duplicar el número de lados tanto en los polígonos inscritos como en los circunscritos, determinación que Arquímedes realizó. De ello extrajo un procedimiento que, iterándolo suficiente número de veces, permite en principio calcular pi con cualquier número de cifras. (Es preciso que pueda calcularse fácilmente el perímetro de un polígono cualquiera con el auxilio de funciones trigonométricas sencillas, como seno, coseno y tangente. Pero en tiempos de Arquímedes, en el siglo III a. de C., se poseía un conocimiento parcial de esas funciones. Arquímedes tuvo, por consiguiente, que fundarse casi exclusivamente en construcciones geométricas, por lo que los cálculos resultaban bastante más duros de lo que hoy pueda parecer.)

Arquímedes empezó inscribiendo y circunscribiendo hexágonos, lo que le proporcionó la desigualdad $3 < \pi < 2\sqrt{3}$. Al duplicar cuatro veces consecutivas el número de lados, hasta 96, estrechó la gama de valores limitantes de pi hasta $3^{10}/_{71}$ y $3^{1}/_{7}$, obteniendo la estimación $\pi \approx 3{,}14$. Parece haber ciertas pruebas de que el texto de *Medición de un círculo* que nos ha llegado constituye sólo un fragmento de otra obra más extensa, en la cual Arquímedes describía cómo, partiendo de decágonos y duplicándolos seis veces, se obtenía una aproximación de cinco cifras: $\pi \approx 3{,}1416$.

El método de Arquímedes resulta conceptualmente sencillo. Ahora bien, si se carece de un método rápido para calcular los valores de las funciones trigonométricas, resulta obligada la extracción de raíces, operación muy lenta y penosa cuando se realiza a mano. Además, las estimaciones así obtenidas convergen lentamente a pi; su error decrece en un factor de alrededor de cuatro en cada iteración. A pesar de ello, todos los intentos de calcular el número pi realizados en Europa hasta mediados del siglo XVII se fundaron de un modo u otro en ese método. Ludolph van Ceulen, matemático holandés del siglo XVI, dedicó gran parte de su carrera al cálculo de pi. Casi al final de su vida obtuvo una aproximación de 32 cifras calculando el perímetro de polígonos inscritos y circunscritos de 2^{62} (unos 10^{18}) lados. Se dice que el valor de pi que obtuvo así, denominado número ludolfiano en ciertas regiones de Europa, fue su epitafio.

El desarrollo del cálculo diferencial, obra en gran parte de Isaac Newton y Gottfried Wilhelm Leibniz, permitió calcular pi de forma mucho más expedita. El cálculo proporciona métodos eficaces de obtener la derivada de una función (la tasa de cambio del valor de la función al cambiar los valores de su variable) y su integral (la suma de los infinitos valores de la función correspondientes a un determinado intervalo de su variable). Aplicando esas técnicas se demuestra que las funciones trigonométricas inversas están dadas mediante integrales de funciones cuadráticas que describen la curva de un círculo. (La inversa de una función trigonométrica da el ángulo correspondiente a cada valor particular de la función. Por ejemplo, el valor de la inversa de la tangente —llamada "arco tangente"— cuando *x* vale 1 es 45 grados o, lo que es igual, $\pi/4$ radianes.)

(Se aprecia la conexión subyacente entre las funciones trigonométricas y

LAS FUNCIONES MODULARES Y LAS APROXIMACIONES DE PI

Una función modular es una función $\lambda(q)$, que puede relacionarse mediante una expresión algebraica, llamada ecuación modular, con la misma función expresada mediante la misma variable q elevada a una potencia entera: $\lambda(q^p)$. El exponente p, que ha de ser entero, determina el "orden" de la ecuación modular. Un ejemplo de función modular es

$$\lambda(q) = 16q \prod_{n=1}^{\infty} \left(\frac{1+q^{2n}}{1+q^{2n-1}}\right)^8 .$$

Su ecuación modular asociada es de séptimo orden. Relaciona $\lambda(q)$ con $\lambda(q^7)$, y está dada por

$$\sqrt[8]{\lambda(q)\lambda(q^7)} + \sqrt[8]{[1-\lambda(q)][1-\lambda(q^7)]} = 1 .$$

Aquellas soluciones de la ecuación modular que satisfacen además otros requisitos especiales reciben el nombre de valores singulares. Una de las clases de valores singulares procede del cálculo de una sucesión de valores, k_p, donde

$$k_p = \sqrt{\lambda(e^{-\pi\sqrt{p}})}$$

en la cual p toma valores enteros. Estos valores tienen la curiosa propiedad de que la expresión logarítmica

$$\frac{-2}{\sqrt{p}} \log\left(\frac{k_p}{4}\right)$$

coincide con muchas de las primeras cifras decimales de pi. El número de cifras que esta expresión tiene en común con pi aumenta al crecer el valor de p.

Ramanujan no tuvo rival en el cálculo de estos valores singulares. Uno de los más famosos es el valor correspondiente al caso $p = 210$, que ya figuraba en su primera carta a G. H. Hardy. Hélo aquí:

$$k_{210} = (\sqrt{2}-1)^2(2-\sqrt{3})(\sqrt{7}-\sqrt{6})^2(8-3\sqrt{7})(\sqrt{10}-3)^2(\sqrt{15}-\sqrt{14})(4-\sqrt{15})^2(6-\sqrt{35}) .$$

Al introducir el valor de este número en la expresión logarítmica, el resultado coincide con las primeras veinte cifras decimales de pi. En comparación con él, k_{240} proporciona un número que coincide con pi a lo largo de más de un millón de cifras.

Aplicando esta técnica general, Ramanujan construyó varias series notables asociadas con el número pi, entre ellas, la de la figura 3. Esa misma técnica general subyace a los algoritmos iterativos, en dos pasos, de la figura 4. El primer paso de cada iteración (el cálculo de y_n), corresponde al cálculo de uno de los términos de una sucesión de valores singulares que se obtiene resolviendo una ecuación modular de orden apropiado. El segundo paso (el cálculo de α_n) equivale a tomar el logaritmo del valor singular.

las expresiones algebraicas considerando un círculo de radio 1 y centro el origen de un plano cartesiano *x-y*. La ecuación de la circunferencia correspondiente, que encierra un área de valor numérico igual a pi, es $x^2 + y^2 = 1$, fórmula que no es más que una reformulación del teorema de Pitágoras correspondiente a un triángulo rectángulo de catetos x e y cuya hipotenusa mida 1. Además, el seno y el coseno del ángulo comprendido entre el semieje x positivo y cualquier punto de la circunferencia son, respectivamente, iguales a la ordenada, y, y a la abscisa, x, de dicho punto; la tangente del ángulo es, sencillamente, y/x.)

No obstante, a los efectos del cálculo de pi, más importancia reviste el hecho de que una función trigonométrica inversa admita un "desarrollo" en serie, cuyos términos son calculables a partir de las derivadas sucesivas de la función. El propio Newton calculó pi con 15 cifras decimales sumando unos cuantos de los primeros términos de una serie que se puede deducir como expresión de la inversa de la función seno. Posteriormente confesó a un colega: "Me da vergüenza confesar a cuántas cifras llevé estos cálculos, que realicé por no tener otra cosa que hacer en aquel momento".

En 1674 Leibniz dedujo la fórmula $1 - 1/3 + 1/5 - 1/7... = \pi/4$, que es el arco tangente de 1. (La serie general del arco tangente la había descubierto el matemático escocés James Gregory. Al parecer, varios siglos antes se desarrollaron en la India expresiones similares.) El error de aproximación, que se define como la diferencia entre la suma de n términos y el valor exacto de $\pi/4$, es sensiblemente igual al $(n + 1)$-ésimo término de la serie. Dado que el denominador de cada uno de los sucesivos términos tan sólo aumenta en 2, es preciso sumar alrededor de 50 términos para lograr una precisión de 2 cifras, 500 términos para disponer de 3, y así sucesivamente. Como es obvio, resulta inabordable dedicarse a sumar los términos de la serie si aspiramos a calcular un valor de pi que supere las pocas cifras.

Sin embargo, una observación realizada por John Machin hizo practicable el cálculo de pi mediante un desarrollo en serie asociado a la función arco tangente. Machin señaló que $\pi/4$ es igual a 4 veces el arco tangente de 1/5 menos el arco tangente de 1/239. Dado que la serie asociada al arco tangente converge tanto más rápidamente cuanto menor es el valor del argumento, la fórmula de Machin simplificó los cálculos. Combinando su fórmula con el desarrollo en serie del arco tangente, Machin calculó en 1706 las cien primeras cifras de pi. En efecto, su técnica se demostró tan potente que todas las evaluaciones posteriores del número pi, desde comienzos del siglo XVIII hasta tiempos recientes, se fundaron en variantes de su método.

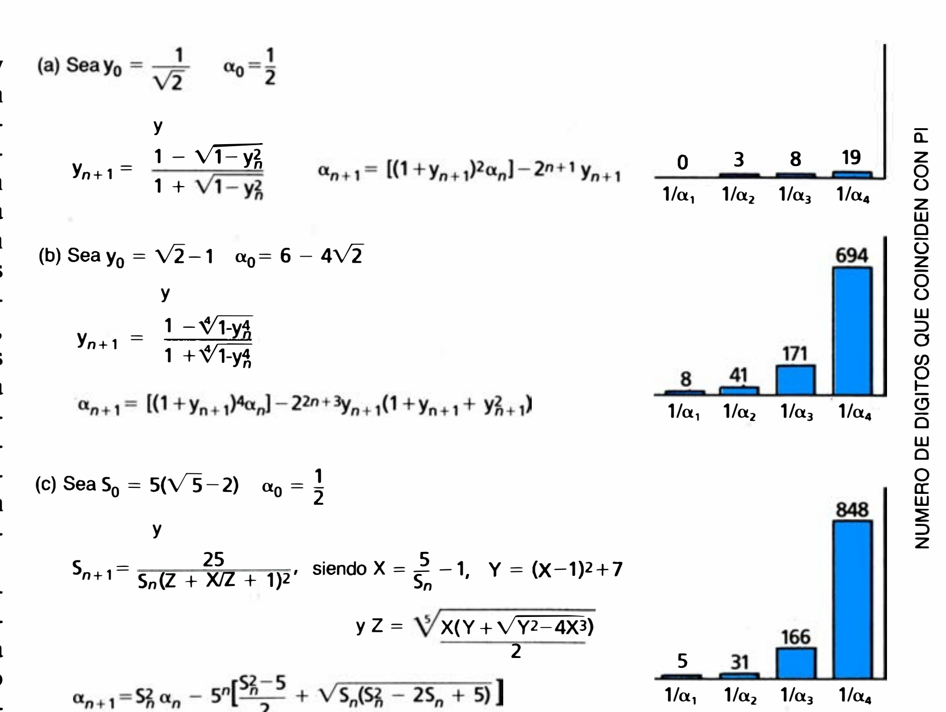
Dos son los cálculos realizados en el siglo XIX dignos de especial mención. En 1844, Johann Dase computó en cosa de meses 205 cifras del número pi por el procedimiento de calcular los valores de tres arcos tangentes valiéndose de una fórmula similar a la de Machin. Dase era un calculador prodigioso, capaz de multiplicar de memoria dos números de 100 cifras, proeza que le llevó aproximadamente ocho horas. (Se diría que fue el precursor inmediato de los modernos superordenadores, al menos en lo que a capacidad de memorización se refiere.) En 1853, William Shanks rebasó de largo a Dase con la publicación del cálculo de pi hasta las 607 cifras, si bien las posteriores a la 527 resultaron ser erróneas. El trabajo de Shanks le llevó muchos años y fue fruto de una aplicación bastante rutinaria, aunque laboriosa, de la fórmula de Machin. (Sí parece constituir una plusmarca que se tardara 92 años en detectar el error de Shanks, que salió a la luz al comparar su valor con una aproximación de 530 cifras obtenida por D. F. Ferguson ayudándose de una calculadora mecánica.)

El advenimiento del ordenador digital trajo consigo un renacer de los esfuerzos por calcular todavía más cifras de pi, pues esas máquinas se avienen de forma ideal a "masticar números" de modo repetitivo. En junio de 1949, John von Neumann y sus colegas aplicaron a la tarea uno de los primeros ordenadores electrónicos, el ENIAC, que generó 2037 dígitos en 70 horas. En 1957, G. E. Felton trató de calcular 10.000 dígitos de pi; mas, por un error de la máquina, sólo resultaron ser correctos los 7480 primeros. La meta de las 10.000 cifras la alcanzó el año siguiente F. Genuys, con un ordenador

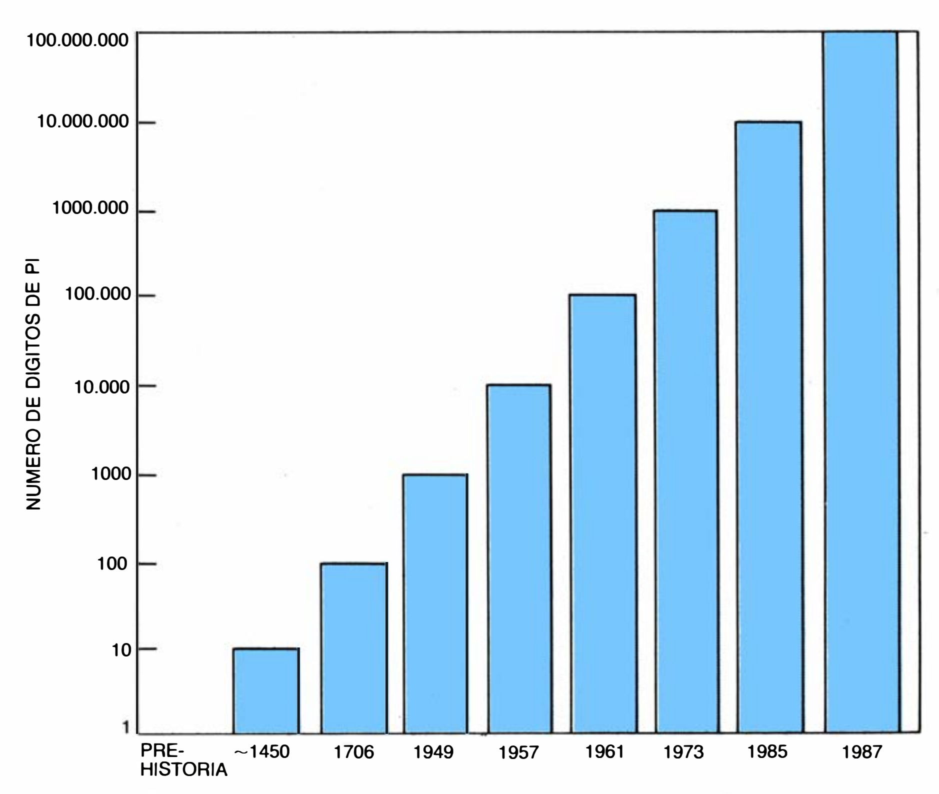
IBM 704. En 1961, Daniel Shanks y John W. Wrench, Jr., calcularon 100.000 cifras de pi en menos de 9 horas con un ordenador IBM 7090. El millón de cifras se rebasó en 1973; Jean Guilloud y M. Bouyer realizaron la proeza, que llevó un poco menos de un día a un CDC 7600. (En realidad, los cálculos de Shanks y Wrench y de Guilloud y Bouyer se realizaron dos veces, usando, para obtener pi, diferentes identidades del arco tangente. En vista del historial de errores en que han incurrido tanto humanos como máquinas, los modernos "cazadores de cifras" no le dan validez oficial a ningún récord hasta que no se realiza tal verificación.)

La causa principal de que se calculara un número cada vez mayor de cifras exactas de pi fue el aumento de la velocidad de las máquinas, pero pronto quedó clara la existencia de límites insuperables. Si se utilizan los métodos tradicionales de realizar operaciones aritméticas en los ordenadores, la duplicación del número de cifras a calcular multiplica como mínimo por un factor 4 el tiempo de cómputo. Por consiguiente, incluso admitiendo que la velocidad de cómputo se multiplicase por 100, el programa de Guilloud y Bouyer hubiera exigido al menos un cuarto de siglo para dar un valor de pi de mil millones de cifras. Desde la perspectiva de los primeros años de la década de 1970, tal computación no parecía abordable.

Sin embargo, hoy resulta factible esa tarea, merced no sólo a que los ordenadores son más rápidos, sino, sobre todo, a métodos nuevos y más eficaces de multiplicar mediante ordenador dos números grandes. Tuvo igual importancia un tercer acontecimiento: el advenimiento de algoritmos iterativos que convergen rápidamente hacia pi. (Los algoritmos iterativos vienen a ser programas de ordenador que efectúan reiteradamente unas mismas operaciones aritméticas, tomando como entrada de cada ciclo la salida del precedente.) Tales algoritmos, algunos de los cuales son obra de los autores, fueron anticipados en muchos aspectos por Ramanujan, a pesar de que nada podía saber de programación de ordenadores. En realidad, los ordenadores no sólo han posibilitado la aplicación de los trabajos de Ramanujan, sino que también han ayudado a descifrarlos. Mediante programas muy refinados de manipulación algebraica ha podido proseguirse la exploración de las sendas que Ra-



4. ALGORITMOS ITERATIVOS (que toman como entrada de cada ciclo la salida del precedente) preparados por los autores. Proporcionan valores muy exactos de pi. (Un algoritmo iterativo consta de una serie finita de operaciones sucesivas que se repiten cíclicamente de tal modo que la salida de cada ciclo constituye la entrada del siguiente.) El algoritmo *a* converge cuadráticamente en $1/\pi$; es decir, el número de cifras correctas proporcionadas por α_n se duplica y más cada vez que *n* crece en una unidad. El algoritmo *b* converge cuárticamente, y el *c* lo hace quínticamente, con lo cual el número de cifras correctas obtenidas tras cada iteración se multiplica, respectivamente, por un factor mayor que cuatro o que cinco. Posiblemente, el algoritmo *b* sea el más eficiente de los conocidos para el cálculo de pi; lo han utilizado los tres últimos cálculos "récord" en superordenador. Durante su análisis de los algoritmos, los autores advirtieron claramente que Ramanujan había seguido métodos similares para obtener sus aproximaciones de pi. De hecho, en el algoritmo *c* el cálculo de s_n se funda en una ecuación modular de quinto orden descubierta por Ramanujan.



5. NUMERO DE DIGITOS conocidos de pi, que se ha incrementado en dos órdenes de magnitud (productos por 10) durante el decenio pasado gracias al desarrollo de algoritmos iterativos capaces de operar en superordenadores equipados con métodos de multiplicación nuevos y eficaces, de velocidad muy superior.

manujan hubo de recorrer en solitario y sin ayuda hace más de 75 años.

Una de las lecciones más interesantes de la informática teórica es que muchos de los algoritmos que nos son familiares, como la regla de multiplicación que se les enseña a los niños en la escuela, distan mucho de ser óptimos. Los teóricos calibran la eficiencia de un algoritmo determinando su complejidad en bits, esto es, el número de veces que se suma o multiplica cada dígito al aplicar el algoritmo. De acuerdo con esta medición, la complejidad de la suma de dos números de n dígitos por el método habitual aumenta en proporción a n y, la multiplicación de dos números de n dígitos, en proporción a n^2. Al aplicar métodos tradicionales, la multiplicación resulta mucho más "penosa" que la adición, en el sentido de que consume un tiempo mucho mayor.

No obstante, como demostraron en 1971 A. Schönhage y V. Strassen, en teoría, la complejidad de la multiplicación de dos números puede ser sólo ligeramente superior a la de la adición. Se logra, por ejemplo, esa reducción potencial de la complejidad aplicando las llamadas "transformadas rápidas de Fourier", o FFT (por *fast Fourier transforms*). La multiplicación de dos números grandes por medio de las transformadas rápidas permite orquestar cuidadosamente los cómputos intermedios entre dígitos y, de ese modo, evitar las redundancias. Dado que la división y la extracción de raíces pueden reducirse a una secuencia de multiplicaciones, también la complejidad, en bits, de estas operaciones puede ser sólo ligeramente mayor que la de la adición. Se logra así un tremendo ahorro de complejidad y, por consiguiente, de tiempo de computación. Por este motivo, todos los esfuerzos recientes en el cálculo de pi se fundan en alguna variante de la técnica de multiplicación mediante la transformada rápida de Fourier.

Empero, para el cálculo real de centenares de millones de dígitos de pi fue preciso redescubrir una preciosa fórmula que ya conocía Carl Friedrich Gauss hace siglo y medio. A mediados del decenio pasado, Richard P. Brent y Eugene Salamin observaron, independientemente, que la fórmula producía un algoritmo para el cálculo de pi cuya convergencia era cuadrática, esto es, era tal que el número de dígitos se duplicaba en cada iteración. A partir de 1983, Yasumasa Kanada y sus colegas, de la Universidad de Tokyo, se han valido de ese algoritmo en el establecimiento de varios récords mundiales en el número de cifras de pi.

Por nuestra parte, nos preguntamos cuál era la razón de la notable convergencia en pi del algoritmo Gauss-Brent-Salamin. Al analizarlo, pusimos a punto técnicas generales para la redacción de similares algoritmos de convergencia rápida, en pi o en otras cantidades. Al trabajar en una teoría esbozada en 1829 por el matemático alemán Karl Gustav Jacobi, advertimos que, en principio, cabía llegar a valores próximos a pi efectuando la evaluación de integrales de una clase llamada "integrales elípticas", que permiten calcular el perímetro de una elipse. (El círculo, fundamento geométrico de los anteriores esfuerzos por calcular aproximadamente el número pi, no es más que una elipse cuyos ejes tienen la misma longitud.)

En general, aunque las integrales elípticas no pueden calcularse directamente por los métodos de integración explícita, sus valores sí se evalúan fácilmente por medio de procedimientos iterativos fundados en las ecuaciones modulares. Hemos hallado que el algoritmo Gauss-Brent-Salamin constituye un caso particular de una técnica nuestra más general, basada en una ecuación modular de segundo orden. Se alcanzaría una convergencia todavía más rápida hacia el valor de la integral y, en consecuencia, un algoritmo más rápido para el cálculo de pi, utilizando ecuaciones modulares de orden superior; por ese motivo, hemos redactado diversos algoritmos basados en ecuaciones modulares de grados tercero, cuarto y superiores.

En enero de 1986, David H. Bailey, del Centro de Investigación Ames, de la NASA, generó 29.360.000 cifras decimales de pi por iteración de uno de nuestros algoritmos en un superordenador Cray-2. Dado que el algoritmo se basa en una ecuación modular de cuarto orden, converge hacia pi cuárticamente, con lo que, en cada iteración, el número de cifras determinadas se multiplica por más de cuatro. Un

METODO DE OBTENCION DE DOS MIL MILLONES DE CIFRAS DE PI CON UNA CALCULADORA*

Sea

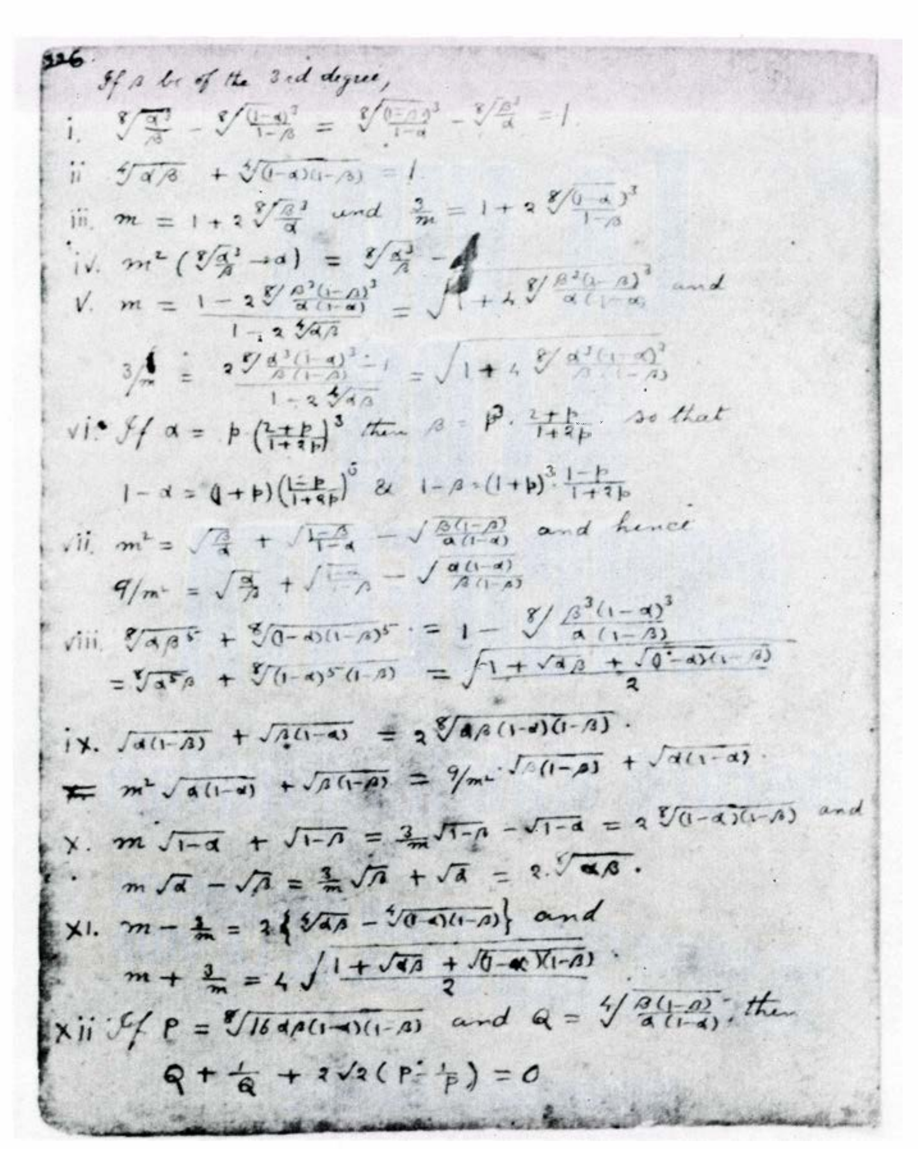
$y_0 = \sqrt{2} - 1$	$\alpha_0 = 6 - 4\sqrt{2}$
$y_1 = [1 - \sqrt[4]{1 - y_0^4}]/[1 + \sqrt[4]{1 - y_0^4}]$	$\alpha_1 = (1 + y_1)^4 \alpha_0 - 2^3 y_1(1 + y_1 + y_1^2)$
$y_2 = [1 - \sqrt[4]{1 - y_1^4}]/[1 + \sqrt[4]{1 - y_1^4}]$	$\alpha_2 = (1 + y_2)^4 \alpha_1 - 2^5 y_2(1 + y_2 + y_2^2)$
$y_3 = [1 - \sqrt[4]{1 - y_2^4}]/[1 + \sqrt[4]{1 - y_2^4}]$	$\alpha_3 = (1 + y_3)^4 \alpha_2 - 2^7 y_3(1 + y_3 + y_3^2)$
$y_4 = [1 - \sqrt[4]{1 - y_3^4}]/[1 + \sqrt[4]{1 - y_3^4}]$	$\alpha_4 = (1 + y_4)^4 \alpha_3 - 2^9 y_4(1 + y_4 + y_4^2)$
$y_5 = [1 - \sqrt[4]{1 - y_4^4}]/[1 + \sqrt[4]{1 - y_4^4}]$	$\alpha_5 = (1 + y_5)^4 \alpha_4 - 2^{11} y_5(1 + y_5 + y_5^2)$
$y_6 = [1 - \sqrt[4]{1 - y_5^4}]/[1 + \sqrt[4]{1 - y_5^4}]$	$\alpha_6 = (1 + y_6)^4 \alpha_5 - 2^{13} y_6(1 + y_6 + y_6^2)$
$y_7 = [1 - \sqrt[4]{1 - y_6^4}]/[1 + \sqrt[4]{1 - y_6^4}]$	$\alpha_7 = (1 + y_7)^4 \alpha_6 - 2^{15} y_7(1 + y_7 + y_7^2)$
$y_8 = [1 - \sqrt[4]{1 - y_7^4}]/[1 + \sqrt[4]{1 - y_7^4}]$	$\alpha_8 = (1 + y_8)^4 \alpha_7 - 2^{17} y_8(1 + y_8 + y_8^2)$
$y_9 = [1 - \sqrt[4]{1 - y_8^4}]/[1 + \sqrt[4]{1 - y_8^4}]$	$\alpha_9 = (1 + y_9)^4 \alpha_8 - 2^{19} y_9(1 + y_9 + y_9^2)$
$y_{10} = [1 - \sqrt[4]{1 - y_9^4}]/[1 + \sqrt[4]{1 - y_9^4}]$	$\alpha_{10} = (1 + y_{10})^4 \alpha_9 - 2^{21} y_{10}(1 + y_{10} + y_{10}^2)$
$y_{11} = [1 - \sqrt[4]{1 - y_{10}^4}]/[1 + \sqrt[4]{1 - y_{10}^4}]$	$\alpha_{11} = (1 + y_{11})^4 \alpha_{10} - 2^{23} y_{11}(1 + y_{11} + y_{11}^2)$
$y_{12} = [1 - \sqrt[4]{1 - y_{11}^4}]/[1 + \sqrt[4]{1 - y_{11}^4}]$	$\alpha_{12} = (1 + y_{12})^4 \alpha_{11} - 2^{25} y_{12}(1 + y_{12} + y_{12}^2)$
$y_{13} = [1 - \sqrt[4]{1 - y_{12}^4}]/[1 + \sqrt[4]{1 - y_{12}^4}]$	$\alpha_{13} = (1 + y_{13})^4 \alpha_{12} - 2^{27} y_{13}(1 + y_{13} + y_{13}^2)$
$y_{14} = [1 - \sqrt[4]{1 - y_{13}^4}]/[1 + \sqrt[4]{1 - y_{13}^4}]$	$\alpha_{14} = (1 + y_{14})^4 \alpha_{13} - 2^{29} y_{14}(1 + y_{14} + y_{14}^2)$
$y_{15} = [1 - \sqrt[4]{1 - y_{14}^4}]/[1 + \sqrt[4]{1 - y_{14}^4}]$	$\alpha_{15} = (1 + y_{15})^4 \alpha_{14} - 2^{31} y_{15}(1 + y_{15} + y_{15}^2)$

$1/\alpha_{15}$ coincide con pi en más de dos mil millones de cifras decimales

*Por supuesto, la calculadora ha de tener una pantalla capaz de presentar dos mil millones de cifras; en una calculadora de bolsillo el cálculo carecería de interés tras la segunda iteración.

6. INSTRUCCIONES EXPLICITAS para ejecutar el algoritmo *b* de la figura 4. Permiten, en teoría, calcular en pocos minutos los dos mil primeros millones de cifras del número pi. Todo cuanto se precisa es una calculadora que disponga de dos registros de memoria y que sea capaz de efectuar las operaciones habituales de suma, resta, multiplicación, división y extracción de raíces. Desafortunadamente, las pantallas de la mayoría de las calculadoras tan sólo muestran ocho cifras, lo cual convierte tal cálculo en pura ficción.

año más tarde, Kanada y sus colegas realizaron una iteración más, alcanzando 134.217.000 cifras decimales en un superordenador NEC SX-2, lo cual les permitió asimismo verificar un cálculo anterior que habían realizado utilizando el algoritmo Gauss-Brent-Salamin. (Iterando nuestro algoritmo un par de veces más —hazaña perfectamente realizable si se nos permitiera monopolizar un superordenador durante unas pocas semanas— se obtendrían más de dos mil millones de cifras del número pi.)



7. "CUADERNOS" DE RAMANUJAN, archivos personales en los que anotaba muchas de sus fórmulas. La página aquí mostrada contiene diversas ecuaciones modulares de tercer orden, todas ellas expresadas en la notación particular que utilizaba Ramanujan. Desdichadamente, el matemático indio no se molestaba en consignar demostraciones formales de sus ecuaciones; otros han tenido que recopilarlas, adaptarlas y demostrarlas. Las fórmulas de los "Cuadernos" constituyen la plasmación de sutiles relaciones entre números y funciones, que pueden aplicarse en múltiples campos de las matemáticas e incluso en la física teórica.

Los métodos iterativos resultan especialmente idóneos para el cálculo de pi mediante ordenadores, pero no a mano; por ello, mal puede sorprender que Ramanujan no se tomase jamás la molestia de efectuarlos. Sí se encuentran en la obra de Ramanujan los ingredientes básicos de los algoritmos iterativos para el cálculo de pi, en particular las ecuaciones modulares. Ciertas partes de su deducción de series y aproximaciones infinitas del número pi tuvieron que correr paralelas a nuestros esfuerzos por obtener algoritmos para pi. En efecto, las fórmulas que menciona en sus "Cuadernos" nos resultaron de gran ayuda en la construcción de algunos de nuestros algoritmos. Por ejemplo, si bien logramos demostrar la existencia de un algoritmo de orden 11, y aunque conocíamos su formulación general, no logramos descubrir su forma inesperadamente simple hasta que tropezamos con las ecuaciones modulares de Ramanujan del mismo orden.

Recíprocamente, hemos logrado deducir la totalidad de las series de Ramanujan a partir de las fórmulas generales desarrolladas por nosotros. La deducción de una de ellas, cuya convergencia hacia pi era más rápida que ninguna de las series que entonces conocíamos, se logró por la ayuda que nos brindó una fuente insospechada. Habíamos justificado la totalidad de las cantidades que figuran en la expresión de la serie, a excepción de una: el coeficiente 1103, que aparece en el numerador de la expresión [*véase la figura 3*]. Estábamos convencidos —como tuvo que estarlo Ramanujan— de que el valor 1103 era correcto. Para demostrar que así era teníamos dos opciones: simplificar una ecuación capaz de amilanar a cualquiera, donde aparecían variables elevadas a potencias de varios millares, o sumergirnos a profundidades considerablemente mayores en los arcanos de la teoría de números.

Por coincidencia, R. William Gosper, Jr., de Symbolics, Inc., había decidido en 1985 sacar partido de esa misma serie de Ramanujan para calcular un valor más preciso de pi que los conocidos. Cuando efectuó el cálculo, llevándolo a más de 17 millones de cifras (la marca del momento), no existía, que él supiera, ninguna demostración de que la serie realmente convergiera hacia pi. Por supuesto, Gosper sabía que millones de cifras del valor que había calculado coincidían con las de un cálculo anteriormente realizado por Kanada mediante el algoritmo Gauss-Brent-Salamin; la posibilidad de error era, pues, infinitesimalmente pequeña.

Sin embargo, en cuanto Gosper completó su cálculo y lo cotejó con el de Kanada, dispusimos de lo necesario para demostrar que el número requerido para hacer que la serie diera valores de pi con error menor que uno entre $10^{10.000.000}$ era 1103. Por un razonamiento muy similar al de que si dos enteros se diferencian en menos de una unidad tienen que ser iguales, su resultado fue suficiente para especificar el número en cuestión: tiene que ser exactamente 1103. De hecho, el cómputo realizado por Gosper pasó a formar parte de nuestra demostración. Sabíamos que la serie (y su algoritmo asociado) era tan sensible a las más leves inexactitudes que, si Gosper hubiera utilizado para el coeficiente cualquier otro valor, o, por otra parte, si el ordenador hubiera introducido una sola cifra errónea en el proceso de cálculo, el resultado hubiera sido un galimatías numérico, en lugar de un valor de pi.

Se demuestra que los algoritmos de tipo Ramanujan para la determinación de valores aproximados de pi se hallan muy cerca de los óptimos posibles. Teniendo en cuenta todas las operaciones

que intervienen en la ejecución de los algoritmos (dando por supuesto que para efectuar la adición, la multiplicación y la extracción de raíces se aplican las mejores técnicas conocidas), la complejidad, en bits, del cálculo de *n* cifras de pi sólo resulta marginalmente mayor que la de multiplicar dos números de *n* cifras. Pero la multiplicación de dos números de *n* cifras mediante la transformada rápida de Fourier es sólo marginalmente más complicada que la adición de dos números de *n* cifras, que es la más sencilla de las operaciones aritméticas posibles en un ordenador.

Es probable que las matemáticas no hayan acusado en toda su plenitud el impacto del genio de Ramanujan. Hay en los "Cuadernos" otras muchas fórmulas maravillosas, que giran en torno a integrales, series infinitas y fracciones continuas (en las que aparece un número más una fracción, cuyo denominador es a su vez expresable como un número más una fracción, y así sucesivamente). Por desdicha, las fórmulas de Ramanujan están dadas sin apenas indicación —si alguna— del método de que se valió para demostrarlas. Littlewood escribió al respecto: "Si en un lugar cualquiera se presentaba un razonamiento importante, y la conjunción total de pruebas e intuición le proporcionaba certeza, no seguía examinándolo".

La hercúlea labor de preparar para publicación los "Cuadernos", iniciada hace 60 años por los analistas británicos G. N. Watson y B. N. Wilson, y que hoy está rematando Bruce Berndt, exige una demostración, una fuente y alguna que otra corrección ocasional a cada uno de los muchos miles de teoremas e identidades en ellos enunciados. Una sola línea de los "Cuadernos" puede fácilmente suscitar muchas páginas de comentarios. Dificulta aún más la tarea la notación, no habitual, en que están escritas las fórmulas, por cuyo motivo gran parte de la obra de Ramanujan no accederá a la comunidad matemática hasta que concluya el proyecto de Berndt.

La capacidad excepcional y única de Ramanujan para trabajar intuitivamente con fórmulas complicadas le permitió plantar semillas en un jardín matemático que (tomando prestada de Freeman Dyson la metáfora) sólo ahora está comenzando a florecer. Juntamente con otros muchos matemáticos, esperamos ansiosos cuáles serán las semillas que germinen los años venideros y embellezcan aún más el jardín.

La Peste Negra

Transmitida por las pulgas que parasitan las ratas, esa enfermedad bacteriana ha afligido a la humanidad durante más de mil años. Se desconocen las razones de su aparición y desaparición cíclicas

Colin McEvedy

Entre Europa, Africa del norte y el Próximo Oriente se alcanzaría, en 1346, una población total cercana a los 100 millones de habitantes. En el curso de unos pocos años, la cuarta parte de ellos murió víctima de una nueva y terrorífica enfermedad que se extendió por aquellos territorios, matando a la mayoría de los que tuvieron la desdicha de padecerla. El mal acabó con el crecimiento poblacional que había marcado la evolución de la sociedad medieval: en apenas cuatro años, entre 1346 y 1352, Europa sufrió la pérdida de unos 20 millones de personas. La enfermedad responsable de tan macabra estadística fue la peste bubónica, a la que, oportunamente, se denominó "Peste Negra", designación que ha conservado a lo largo de la historia.

Los efectos de la Peste Negra resultaron particularmente catastróficos, al atacar después de un largo período en el que la enfermedad había estado ausente de Occidente. La que comentamos no fue la primera ocasión en que la peste asoló Europa. Unos 800 años antes, durante el reinado de Justiniano, en el siglo VI, se registró una epidemia de proporciones similares. Durante los dos siglos que siguieron a la peste de Justiniano, como durante los cuatro que siguieron a la Peste Negra, se desencadenaron otras epidemias, aunque de menor entidad. Desde aquellos tiempos la enfermedad ha sufrido un pronunciado declive, si bien todavía hoy se da en algunos lugares, incluso en Estados Unidos.

Entre el 70 y el 80 por ciento de los que contrajeron la peste en el siglo XIV murieron por ello. Los síntomas se presentaban con tal gravedad que presagiaban la muerte en un plazo de cinco días. El nombre de peste bubónica procede de uno de los primeros signos clínicos de la enfermedad: la aparición de grandes hinchazones dolorosas (llamadas bubones) en los nódulos linfáticos de la axila, cuello o ingle de la víctima. Por regla general, tres días después de la aparición de los bubones los enfermos se veían asaltados por fiebre alta, acompañada de delirio y de la presencia de manchas negras, resultado de hemorragias subcutáneas. Conforme avanzaba el proceso, los bubones crecían y se hacían más dolorosos; a menudo estallaban.

Este último episodio resultaba especialmente atroz, hasta el punto de que los pacientes, moribundos ya, se sumían en un estado de gran agitación. Sin embargo, los médicos siempre interpretaron que el estallido constituía una buena señal, aunque sólo fuera porque probaba que los pacientes seguían luchando casi una semana después de iniciado el mal. Posiblemente la mitad de los condenados a morir ya había fallecido antes de alcanzar este estadio.

Los que padecían la infección directa del torrente sanguíneo presentaban un shock séptico caracterizado por hemorragias masivas y muerte rápida, forma de peste conocida como septicémica. En otros casos, la peste se contagiaba como una especie de neumonía; las víctimas sufrían colapsos, escupían sangre y en muy pocos días estaban prácticamente muertas; tal era la peste neumónica.

Por extraño que parezca, pese a la frecuente presencia de la enfermedad y el elevado precio que se cobraba, nadie disponía entonces de indicios sobre su naturaleza, su causa última o el mecanismo de difusión. Durante el período de la Peste Negra, las gentes solían atribuir el mal a combinaciones astrológicas desfavorables o aires o vapores pútridos ("miasmas"), nada de lo cual podía traducirse en programas preventivos de ninguna clase. Algunas teorizaciones paranoides achacaron la enfermedad a un envenenamiento pro-



1. HORRORES CONSECUTIVOS A LA PESTE representados en "El triunfo de la muerte", de Pieter Brueghel el Viejo. La parca, en forma de esqueletos deambulantes, arrasa el reino de los vivos. Ni el rey, con sus toneles de oro, ni los jóvenes juerguistas de la mesa escapan al inexorable ejército de los muertos. Detrás del monarca, una patrulla de esqueletos arroja sus víctimas a una fosa común; en el trasfondo se observa un paisaje desolado, desprovisto ya de vida. Este tipo de visiones apocalípticas eran frecuentes durante los siglos en que la peste devastaba Europa y los más sanos podían desaparecer en cuestión de días.

vocado voluntariamente por brujas, por musulmanes (decían los cristianos), por cristianos (según los musulmanes) o por judíos (según aquellos otros dos grupos).

Por fin, en 1894, el bacteriólogo francés Alexandre Yersin descubrió que la peste bubónica estaba causada por una bacteria gram-negativa, la *Yersinia pestis*, perteneciente al grupo bacteriano de los bacilos filamentosos, muchos de los cuales son patógenos. Los bacilos de la peste se encuentran con baja frecuencia en numerosas poblaciones de roedores salvajes de todo el mundo; se transmiten entre ellos por las pulgas. En el caso de la peste bubónica, el insecto responsable suele ser la pulga oriental de rata, *Xenopsylla cheopis*. Cuando pica a una rata infectada, ingiere los bacilos, que proliferan en su tubo digestivo y lo obstruyen; la pulga, incapaz de ingerir sangre, se vuelve rabiosamente hambrienta y, en sus frenéticos intentos por alimentarse, pica repetidamente al hospedador, regurgitando más bacilos en cada picadura. Los puntos de inyección actúan como focos de diseminación de los bacilos. Si el hospedador muere, cosa probable, la pulga se traslada al roedor más cercano. De este modo la enfermedad se propaga con rapidez. Conforme disminuye el número de ratas vivas, las pulgas parasitan hospedadores de sangre caliente a los que normalmente no atacarían, como seres humanos y animales domésticos, y se desencadena una epidemia.

Una vez introducida la enfermedad entre la población humana, puede transmitirse de una persona a otra a través de la inhalación de gotitas infectadas expulsadas con la respiración. Eso no quita que la manera habitual de propagación sea la picadura de pulga. De hecho, la enfermedad no subsiste al acabarse los roedores, que son los portadores primarios del bacilo de la peste y de la pulga de los múridos.

La condición esencial para que se produzca una epidemia (un brote en la población humana) es la existencia de una epizootia (un brote en la población animal, en este caso de roedores). Y ello es necesario no tanto para iniciar cuanto para mantener la enfermedad entre los humanos. Por supuesto, ambas poblaciones han de encontrarse en íntimo contacto para que la trasmisión tenga éxito, aunque es dudoso que eso resultara nunca un problema en el Medievo. Las personas vivían entonces rodeadas de ratas, tanto en las áreas urbanas como en las rurales.

Se supone que la Peste Negra recorrió la Ruta de la Seda, la senda transasiática utilizada para el transporte de la seda china hasta Europa. Dos razones lo avalan. En primer lugar, en 1346 se registraron brotes de peste en Astrakán y Saray, estaciones de caravanas del bajo Volga, en territorios de lo que hoy es la Unión Soviética. En segundo lugar, durante los años 1347 y 1348, el erudito y viajero árabe Ibn Battuta, de regreso de la India por la Ruta de las Especias, indicó que las primeras noticias sobre la peste las encontró a su llegada a Aleppo, en el norte de Siria, y no antes. Ello excluye la propagación pestífera por la vía del océano Indico y puertos del golfo Pérsico.



Lo más probable es que la enfermedad se iniciara entre las marmotas, grandes roedores originarios del Asia central cuya piel constituía un importante artículo comercial en aquella parte del mundo. Según esta reconstrucción histórica, los tramperos probablemente encontraron animales muertos o moribundos y recogieron las pieles, encantados de hallar tan abundante suministro, vendiéndolas a comerciantes quienes, a su vez, sin dar importancia a las noticias sobre aparición de enfermedades entre los cazadores, las venderían a compradores occidentales. Cuando los fardos de pieles de marmota enviados a través de la Ruta de la Seda se abrieron en Astrakán, las pulgas, hambrientas, saltarían de las pieles en busca de la ración de sangre que tuvieran más a mano. Según parece, desde Saray la peste se trasladó río Don abajo hasta Caffa, un importante puerto del mar Negro, cuya gran población ratuna brindaría el caldo de cultivo ideal para los bacilos pestíferos. Dado que muchas de las ratas de Caffa vivían en barcos veleros que efectuaban viajes a otros puntos de Europa, la enfermedad encontró un fácil medio de transporte hacia esta parte del mundo.

En efecto, difícilmente podría diseñarse un medio más eficaz de diseminar la peste que un barco medieval. Las bodegas de aquellas naves estaban abarrotadas de ratas; cuando la tripulación dormía, las ratas se enseñoreaban del buque, corriendo por las jarcias y sembrando la cubierta de pulgas. El ciclo infeccioso de pulga a rata y viceversa se mantendría hasta que la población ratuna se viera tan disminuida por la enfermedad que no pudiera mantener a todas las pulgas y los consiguientes bacilos pestíferos. Las pulgas, hambrientas, transmitirían entonces el mal a la población humana. No debe extrañar, pues, que, a finales de 1347, la peste afectara a la mayoría de los puertos situados en la ruta que unía Caffa con la ciudad italiana de Génova.



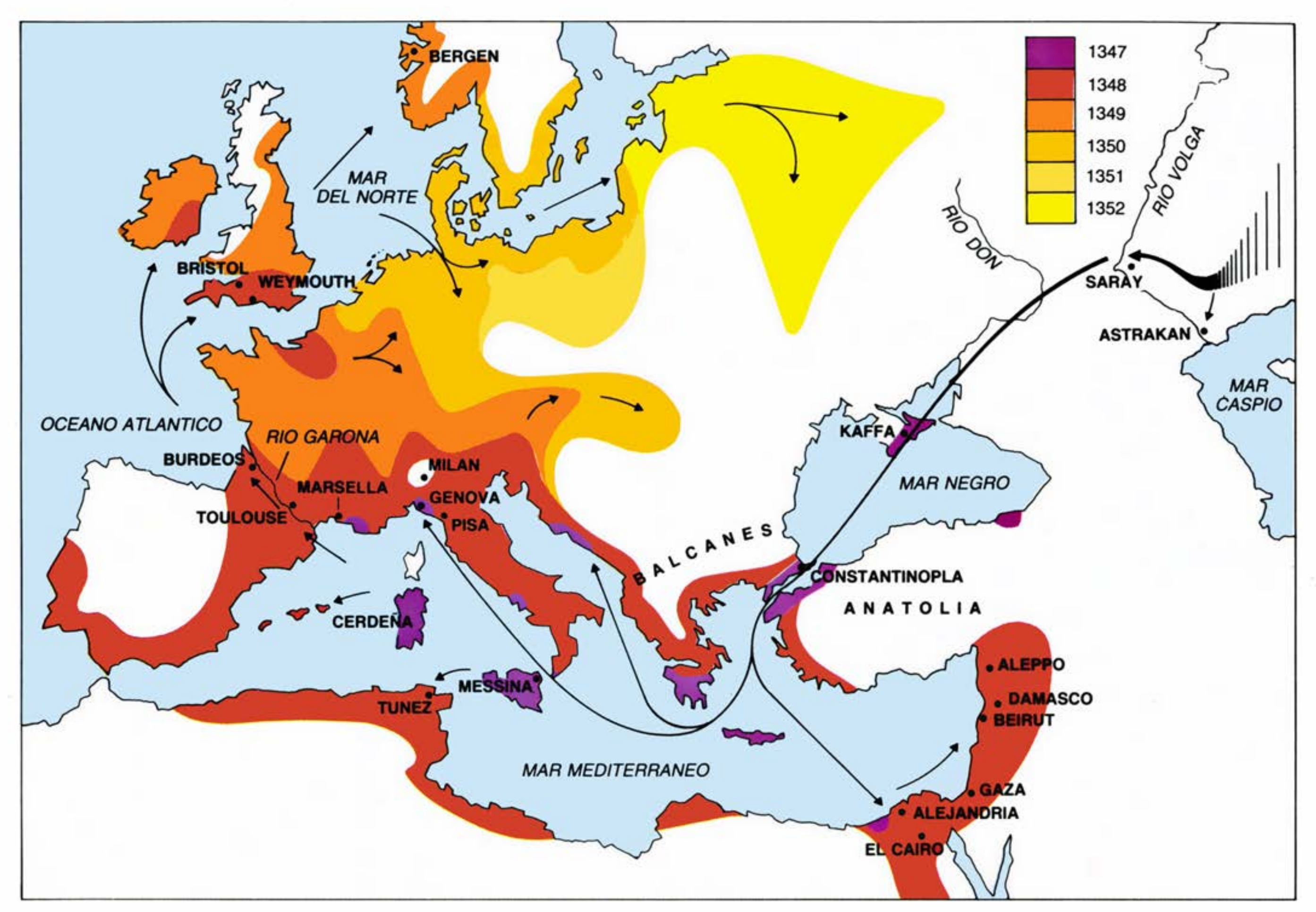
2. TRAJE USADO POR LOS MEDICOS durante el brote de peste que golpeó Marsella en 1720. Recubría a su portador de pies a cabeza; a ésta se le daba apariencia de pájaro, creyéndose que protegía del contagio. En el pico se depositaban hierbas aromáticas a modo de filtro del aire contagiado; la vara contenía incienso para despejar los miasmas. Incluso los orificios de los ojos ofrecían protección, recubiertos de cristal.

Los dos puertos más importantes a lo largo de esa ruta eran Pera, suburbio de Constantinopla, y Messina, en Sicilia. Ambos constituían puntos obligados de atraque de las naves que cruzaban el Mediterráneo y se convirtieron en importantes focos de diseminación. El primer impacto sobre la población de Constantinopla lo describió gráficamente el emperador Juan VI Cantacuzeno, quien perdió un hijo por causa del mal en 1347. Contaba que se extendía por las islas griegas y por las costas de Anatolia y de los Balcanes, matando "a la mayoría de la gente". El primer brote registrado en Messina ocurrió en octubre de 1347, desencadenándose una epidemia que se extendió con rapidez por toda la isla.

Desde allí, a comienzos de 1348, la Peste Negra afectó a Tunicia, en la costa septentrional de Africa, y después se extendió por Cerdeña hasta España. Al tiempo de alcanzar la península Ibérica, la peste bubónica ya se había extendido al corazón de Europa, de lo que puede acusarse, en parte, a los genoveses. En efecto, se cuenta de ellos que impidieron entrar en su puerto naves procedentes de Oriente y cargadas con compatriotas enfermos. Esta actitud insolidaria tuvo escaso éxito, pues la ciudad resultó tan afectada como cualquier otra, y el desvío de los barcos hacia otros puertos, como Marsella y Pisa, aceleró la invasión de Europa por la peste.

Para entonces la peste ya asolaba todo el Mediterráneo. Navíos de transporte de seda, pieles y esclavos la habían transmitido a Alejandría antes de concluir 1347; desde allí se extendió al Sur hacia El Cairo, al Este hacia Gaza, Beirut y Damasco y, finalmente, a lo largo de la costa africana hacia Marruecos.

En 1348 la Peste Negra había saltado



3. LA PESTE NEGRA alcanzó Europa, procedente del Asia Central, a través de la Ruta de la Seda, presentándose en Caffa en torno a 1347. Desde este importante puerto del Mar Negro asaltó las ciudades costeras más importantes de Europa y del norte de Africa. La mayor parte de Europa resultó afectada antes de que la epidemia llegara a su fin, en 1352. Milán fue la más importante de las ciudades que se salvaron, a causa, según se cree, de su lejanía del mar.

del Mediterráneo a la costa atlántica europea. Cruzó el suroeste de Francia a través de su capital, Toulouse, y rápidamente recorrió el Garona hasta Burdeos, en la costa occidental. Desde allí es presumible que uno de los barcos que transportaban vino para Inglaterra la introdujera en Gran Bretaña. En 1348 se registró el primer brote en Weymouth, en la costa sur inglesa, y se extendió después a Irlanda, posiblemente desde Bristol.

Desde Inglaterra, la peste cruzó el mar del Norte para atrapar Escandinavia entre sus mortíferas garras. Según un relato, la invasión de Escandinavia pudo achacarse a un barco que abandonó Londres en mayo de 1349, llevando una tripulación completa y un cargamento de lana en dirección a Bergen. Días después apareció la nave a la deriva frente a las costas noruegas. Gente de la costa, que se acercó a remo, encontró la tripulación muerta; volvieron a tierra con la lana e, involuntariamente, también con la peste. Así empezó una reacción en cadena que afectó, pueblo tras pueblo, la costa noruega.

Durante el año siguiente la Peste Negra hizo estragos entre las poblaciones de Dinamarca y Alemania, antes de entrar en Polonia (1351) y Rusia (1352). De este modo se completó el círculo; no sólo volvía la enfermedad a lugares alejados apenas unos centenares de kilómetros del lugar por donde entró en Europa sino que, después de cuatro años devastadores, las tasas de mortalidad europeas recuperaron los valores normales.

La sociedad que sobrevivió a la Peste Negra conoció una notable prosperidad; los supervivientes heredaron las fortunas de sus familiares fallecidos y muchos ocuparon posiciones prominentes que antes les habían estado vedadas. Sin embargo, su buena suerte no duró mucho. En 1356 apareció un segundo brote de peste bubónica en Alemania, que se extendió con rapidez. El mayor tributo hubieron de pagarlo los niños nacidos con posterioridad a la primera plaga.

Desde entonces, la peste siguió visitando Europa con triste regularidad; el continente no lograba liberarse de ella más que durante períodos de unos pocos años seguidos. Si bien las epidemias siguientes no alcanzaron la mortalidad de la Peste Negra, persistió su incidencia negativa sobre el crecimiento de la población europea hasta finales del siglo XIV.

Se alcanzó entonces el equilibrio entre la peste y los humanos y en el siglo XV la población comenzó a recuperarse. En las regiones fuertemente castigadas, el número de habitantes tardó en volver a sus niveles originales; a finales del XVI, la población superaba ya en todos lados las cifras anteriores a la peste.

Curiosamente, las posteriores reapariciones de la peste (menos frecuentes) alcanzaron, en muchas ocasiones, la virulencia de los peores brotes. Así, durante la última epidemia francesa, entre 1720 y 1722, murió la mitad de la población de Marsella, el 60 por ciento en la vecina Tolón, el 44 por ciento en Arlés y el 30 por ciento en Avignon y Aix. Sin embargo, la epidemia se constriñó a la región de Provenza y el nú-

mero total de fallecidos no superó los 100.000.

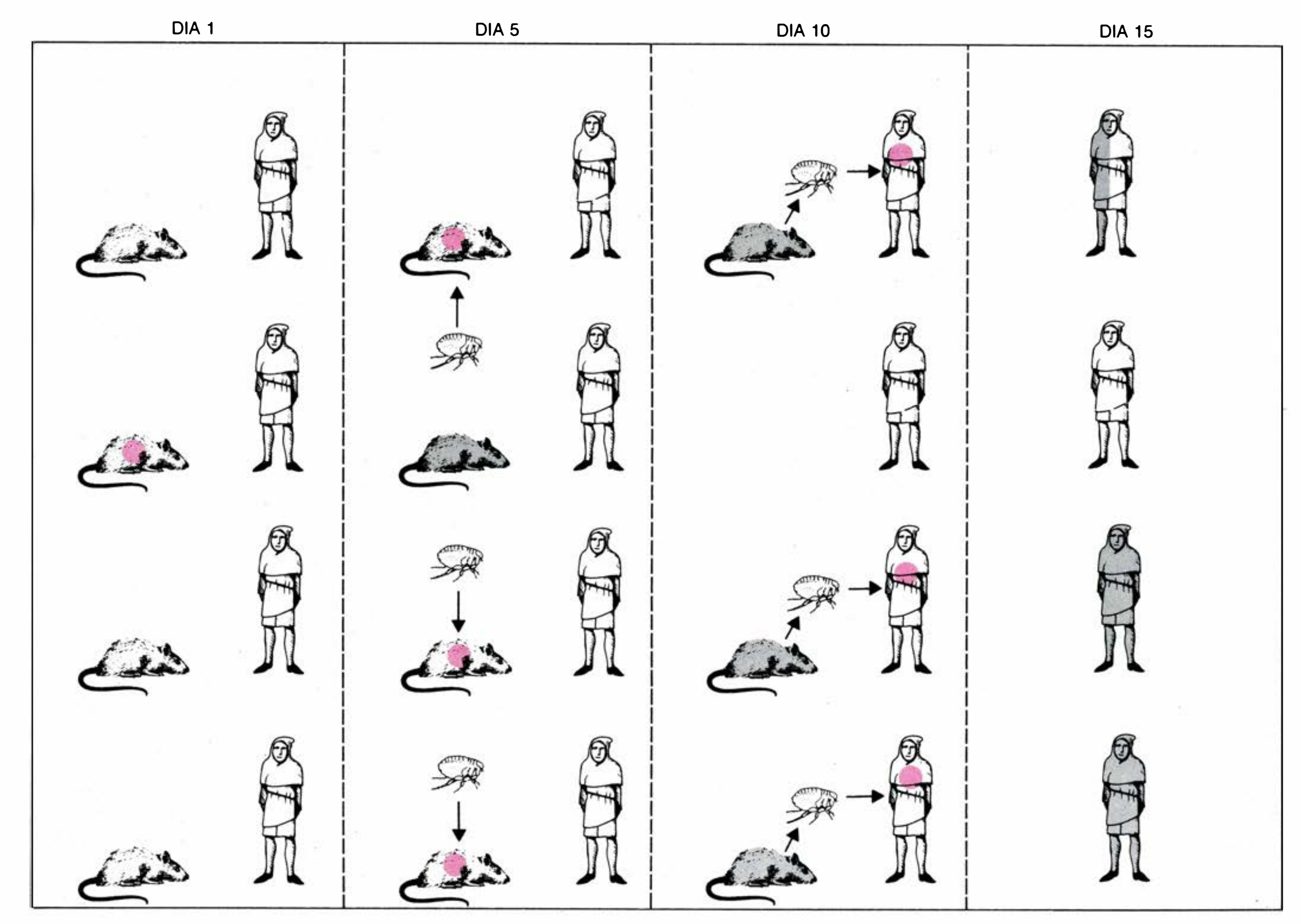
En el siglo XVI era creencia generalizada que la peste se extendía por contagio, es decir, que un factor tóxico se transmitía de los enfermos a los sanos. La transmisión podría ser por contacto directo con un enfermo o, de modo indirecto, a través de sus vestidos o ropas de cama. De acuerdo con ello, numerosas ciudades y pueblos establecieron regulaciones cuarentenarias. Las autoridades inglesas, por ejemplo, recomendaron que los enfermos de peste se encerraran en sus casas o se trasladaran a "casas de apestados" especialmente dispuestas al efecto. Un caso extremo de seguimiento de tales disposiciones fue el famoso protagonizado por William Mompeson, párroco del pueblecito de Gyam, Derbyshire, quien, cuando apareció la peste de 1666, convenció a la comunidad entera para que se encerrara y así pasara la cuarentena. Uno por uno, los feligreses que permanecieron fieles a sus contaminados hogares padecieron la enfermedad. La tasa de mortalidad, del 72 por ciento, indica que la comunidad probablemente sufrió una tasa de morbilidad (infección) del 100 por ciento, precio sin duda extraordinario por una teoría equivocada.

Encerrar la gente en casa es, obviamente, uno de los peores métodos de lucha contra la peste. Esta enfermedad está ligada al local, manifestándose con mayor facilidad cuando las ratas, las pulgas y las personas se mantienen en estrecho contacto mutuo. Enclaustrar a las personas significa aumentar la probabilidad de que les pique una pulga portadora de peste o se infecten por contacto con otro ser humano.

Las autoridades reconocieron que las cuarentenas eran peligrosas para las personas sanas que se encerraban con sus familiares enfermos, pero las mantenían creyendo que, pese al sacrificio de algunas vidas, quizá se frenara la extensión de la enfermedad. Dado que la peste la transportan las ratas, que se movían con libertad, el esfuerzo cuarentenario supuso un pérdida de tiempo y de vidas.

También se hicieron intentos de aplicar la cuarentena a los pasajeros y mercancías transportados en barco. Cuando se observaba la presencia de enfermedades parecidas a la peste entre la tripulación o el pasaje, los barcos se enviaban a lazaretos (estaciones cuarentenarias) hasta que las autoridades consideraban que había pasado el peligro y los liberaban. Por ejemplo, en Marsella, en mayo de 1720, el velero *Grand Saint Antoine* tuvo que pasar una cuarentena de tres semanas porque ocho de sus tripulantes habían fallecido durante la travesía desde el Próximo Oriente. Pese a esos esfuerzos para contener la expansión de la peste, la enfermedad brotó en Marsella, iniciándose entre los trabajadores portuarios que descargaron el buque a su salida de la cuarentena.

Pocas son las pruebas de que ese tipo



4. DISEMINACION DE LA PESTE en una casa medieval; podía ser muy rápida a partir del momento en que se infectaban las ratas negras que la habitaban. Una rata infectada, como la marcada con un punto de color el día 1, moriría el día 5. Entonces sus pulgas la abandonarían, llevando la peste a otras ratas. El día 10 dichos animales también habrían muerto y las pulgas afectarían a los seres humanos, de los que infectarían alrededor del 75 por ciento. El día 15, aproximadamente la mitad de las personas habría muerto, una cuarta parte se habría repuesto de la enfermedad y otra cuarta parte habría escapado indemne.

de cuarentenas resultaran alguna vez eficaces. Venecia fue uno de los primeros puertos que puso en vigor prescripciones de aislamiento cuarentenario, a comienzos del siglo xv, que se hacían cumplir bajo amenaza de pena de muerte para los contraventores. Y sin embargo, Venecia sufrió la peste tanto como cualquier otra ciudad de Italia, presumiblemente porque resultaría imposible evitar que las ratas del buque sometido a cuarentena saltaran a tierra y llevaran la peste con ellas.

Por fin, tras incontables ciclos de presencia y ausencia, la peste desapareció de Europa. El último episodio, que afectó a Londres, comenzó en 1665 y acabó, espectacularmente, en el gran incendio de 1666. Los londinenses, como es natural, atribuyeron su salvación al fuego purificador. Más tarde se sugirió que su resistencia a la peste se debió a la reconstrucción que siguió al fuego y al hecho de que la ciudad nueva dipusiera de casas de ladrillo y calles amplias y limpias, en lugar de las retorcidas estructuras de madera y paja y malolientes callejones de la época medieval.

Es una explicación atrayente que no resiste un examen serio. Por un lado, el fuego sólo destruyó el centro de Londres, el área menos afectada por cualquiera de los anteriores brotes de peste, dejando indemnes los congestionados suburbios que habían proporcionado el principal asiento del mal en las ocasiones anteriores. Por otro lado, otras ciudades europeas, como París y Amsterdam, se libraron de la peste durante la misma época, lo que no puede ponerse en relación con el incendio de Londres.

Otra hipótesis, más sugerente aunque también inexacta, explica la desaparición de la peste por coincidencia con ciertas mejoras en los niveles de higiene y sanidad. Si bien no puede desecharse el factor higiénico, no explica por sí solo que los brotes pestíferos posteriores ocurrieran cada vez más lejos del centro de Europa y mantuvieran su curso habitual y sus elevadas tasas de mortalidad. Parecía como si Europa estuviera desarrollando cierta forma de resistencia a la peste, que impedía su propagación al modo usual. En el Norte, la tendencia a la retirada viajó hacia el Este; en las zonas mediterráneas se dirigió hacia el Sur. Cuanto más tardía era la presentación de la epidemia, menos capacidad de difusión parecía tener. Para mayor admiración, ello ocurría cuando, según todos los índices disponibles, el tráfico por mar y tierra se incrementaba.

Cuando, a últimos del siglo xix, se comprendió por fin el papel desempeñado por las ratas, se sugirió que el apaciguamiento de la peste podría explicarse por cambios de la dinámica de las poblaciones de rata negra, *Rattus rattus*. Durante el siglo xviii se observó que la rata negra, portadora histórica de la enfermedad, había sido desplazada en buena medida por una nueva especie, la rata marrón (*Rattus norvegicus*), que quizá constituyera un vector mucho menos apropiado, puesto que, aun siendo similar su susceptibilidad al bacilo, la rata marrón suele vivir más alejada de los humanos que la negra. Residencias típicas de la rata marrón son los sótanos oscuros, o las alcantarillas, mientras que la rata negra corretea por las habitaciones superiores y por las vigas de la casa. Dado que el salto de la pulga oriental de rata alcanza a lo sumo 90 milímetros de longitud, la diferencia de hábitos pudo bastar para aislar a los humanos de las pulgas portadoras de la peste.

La teoría de la rata marrón parece verosímil, pero no concuerda con los datos geográficos: la rata marrón se extendió por Europa en el siglo xviii, de Este a Oeste, mientras que la peste se retiraba de Oeste a Este. La rata marrón existía en Moscú desde mucho antes que la ciudad sufriera la epidemia, particularmente dura, de los años posteriores a 1770, mientras que no alcanzó Inglaterra hasta 1727, sesenta años después del último brote de peste registrado en aquel país.

Andrew B. Appleby, de la Universidad estatal de San Diego, sugirió una hipótesis alternativa, a saber, que cierto porcentaje de ratas negras adquirieron resistencia a la peste a lo largo del siglo xvii, proliferaron y se extendieron por Europa durante la siguiente centuria. Aunque esas ratas fueran infectadas por el bacilo pestífero, no morirían y, de esa forma, acogerían a una gran población de pulgas, que no tendría necesidad de buscar otros huéspedes. La hipótesis, sin embargo, no se ajusta a lo que sabemos acerca de la resistencia a la peste en poblaciones animales. Como ha indicado Paul Slack, de la Universidad de Oxford, las poblaciones ratunas suelen desarrollar resistencia al exponerlas a una bacteria o virus patógenos, pero tales resistencias desaparecen pronto, por lo que es muy poco probable que el fenómeno explique una inmunidad generalizada frente a la peste.

Otra hipótesis más ajustada propone la aparición de una nueva especie de *Yersinia pestis*, el bacilo pestífero, evolucionada en sentido menos virulento que la estirpe anterior. Al ser menos dañina, habría actuado como una vacuna, confiriendo a los animales y a las personas infectadas una relativa inmunidad frente a las estirpes más virulentas de la misma bacteria.

La hipótesis bacteriológica resulta aceptable por varias razones. Primero, está de acuerdo con el aforismo propuesto por el patólogo norteamericano Theobald Smith, que dice: "las manifestaciones patológicas sólo constituyen incidentes en el desarrollo de un proceso de parasitación"; esto es, que, a largo plazo, las formas más suaves de enfermedad tienden a desplazar a las más virulentas. Segundo, explica por qué la progresiva disminución de la peste coincide con la limitación local de los brotes: una enfermedad no puede propagarse cuando el número de seres susceptibles a ella es bajo. Tercero, viene apoyada por la existencia de un bacilo de parentesco muy próximo al de la peste, *Yersinia pseudotuberculosis*, que no induce manifestaciones visibles de la enfermedad en ratas pero les confiere un elevado grado de inmunidad frente a la peste.

¿Se extendió *Y. pseudotuberculosis*, u otro bacilo semejante, de similares propiedades, entre la población roedora de la Europa moderna, impidiendo que *Y. pestis* se asentara en ella? Pese a que no existen pruebas directas en favor de esta hipótesis, es la que parece más razonable.

El descubrimiento de los antibióticos y su generalizado uso han proporcionado al ser humano otra forma de protección frente a la peste. Aunque la enfermedad sigue presentándose con regularidad en zonas de Africa, Sudamérica y el Suroeste de los Estados Unidos (donde en 1986 se declararon 10 casos), es difícil que vuelva a alcanzar niveles pandémicos, pues sabemos cómo se difunde, cuáles son las medidas sanitarias convenientes y cómo tratar a los enfermos. Muchas otras cuestiones sobre la peste continúan, pese a todo, sin tener solución. Por ejemplo, no está claro el modo de transmisión en las zonas rurales, donde la población ratuna es discontinua. Y, ¿qué explica la actual distribución de la peste en el mundo? ¿Por qué sólo ciertas poblaciones de roedores son depositarias de la peste, mientras que otras están absolutamente libres de ella?

Juegos de ordenador

Un relojero ciego supervisa el país de los biomorfos

A. K. Dewdney

Si se estableciera un paralelismo entre los mecanismos biológicos y los de relojería, forzoso sería concluir que el relojero es ciego. Así lo mantiene Richard Dawkins, biólogo de la Universidad de Oxford, autor de *El gen egoísta*, publicado por la editorial Labor de Barcelona. Después de todo, las leyes que rigen la evolución son fuerzas físicas, ciegas. Dawkins se ha incorporado al debate entre creacionistas y evolucionistas, tomando partido por estos últimos en una obra reciente que titula *El relojero ciego* (que la misma editorial Labor tiene en prensa). Para ilustrar una de las tesis centrales del libro, Dawkins ha preparado un programa que permite remedar el proceso evolutivo y diseñar formas de vida gráfica, criaturas abstractas a las que Dawkins llama biomorfos.

El programa, que denominaré simplemente RELOJERO, arroja luz sobre una falacia que suelen aducir quienes critican la teoría de la evolución, a saber, ¿cómo es posible que algo tan complejo como la maquinaria de la vida haya surgido por combinación aleatoria de acontecimientos? A primera vista, ello es palmariamente imposible. Seguro que buscaríamos un relojero, como hiciera William Paley, teólogo inglés del siglo XVIII. En una argumentación en favor de la existencia de un Creador de todos los seres vivos, Paley se imaginaba tropezando con una piedra al atravesar un baldío. Se diría que siempre había yacido allí. "Pero imaginemos, proseguía Paley, que en el suelo hubiera encontrado un reloj, y que me preguntara también qué hacía el reloj en aquel lugar; mal hubiera pensado en la respuesta que antes dí ... el reloj forzosamente ha debido tener un constructor."

Paley murió antes de que naciera Charles Darwin. Pese a ello, los creacionistas siguen estando de acuerdo con su razonamiento. ¿Gracias a qué combinación de acontecimientos pudo haberse ensamblado un reloj en el baldío de Paley? La falacia estriba en suponer que los acontecimientos aleatorios que han dado por fruto el reloj se registraron simultáneamente, poco más o menos. El razonamiento no toma en consideración los efectos de los cambios acumulativos. RELOJERO es un remedo computacional de la potencia de los cambios acumulativos. Sus productos no son seres vivos, sino biomorfos, esto es, formas estáticas que pueden semejar seres vivos.

Se comienza con una forma muy sencilla situada en el centro de una pantalla dividida en grandes cuadrados [*véase la figura 2*]. Supongamos, por ejemplo, que el cuadrado central contenga un arbolito provisto de unas pocas ramas. RELOJERO produce variantes que ocupan los cuadros adyacentes. Algunos de ellos tendrán más ramas; otros, menos. Las posibles formas de variación de los árboles están controladas por cierto número de genes. Volveré a referirme a ellos más adelante. Todas esas variedades de árboles se consideran progenie del árbol central. Representan las variantes de que dispone la población actual.

RELOJERO funciona en el ordenador Macintosh. En consecuencia, la evolución se halla controlada por un "ratón", el ratón del Macintosh, claro está. El ratón es un dispositivo de entrada que el usuario arrastra manualmente sobre una superficie lisa. El movimiento del ratón sobre tal superficie lo reproduce un cuadradito oscuro que se mueve en perfecta correspondencia sobre la pantalla del Macintosh. Durante el funcionamiento de RELOJERO, el usuario selecciona cuál será el árbol que nazca a continuación desplazando el cursor controlado por el ratón al interior de uno de los cuadros de variante contiguos. El ratón va provisto de un botón y, pulsándolo cuando el cuadrado se halla en el recuadro elegido, el usuario hace que tal árbol se convierta en progenitor de todos los subsiguientes. El árbol elegido migra entonces al recuadro central; el ciclo de selección recomienza. Una secuencia de selecciones genera una línea filogenética de árboles, cada uno de los cuales se diferencia ligeramente de su progenitor. La lenta y regular acumulación de pequeños cambios provoca resultados sorprendentes.

Hallándose a punto de hacer funcionar RELOJERO por primera vez, escribe Dawkins: "Esperaba encontrarme sauces llorones, cedros del Líbano, álamos de Lombardía, algas marinas, astas de ciervo, tal vez. Ni mi intuición de biólogo, mi experiencia de 20 años de programación de ordenadores o mis sueños más descabellados me habían preparado para lo que en realidad salió a la luz en la pantalla. No recuerdo exactamente en qué momento de la secuencia comenzó a abrirse paso en mí la idea de que ... parecía un insecto... Mi incredulidad crecía a la par que se acentuaba la semejanza." Al sonido de triunfales acordes sinfónicos emergieron criaturas claramente similares a insectos [*véase la figura 3*]. Dawkins tuvo aquella noche un sueño inquieto, llenos los párpados de enjambres de hexápodos.

Así surgieron los primeros ejemplares de las formas a las que Dawkins aplicaría el término genérico de biomorfos. En realidad, la totalidad de las formas producidas por RELOJERO no son más que árboles. El efecto combinado de las ramas que se repliegan sobre el tronco, o que se retuercen inesperadamente, no sólo produce cuerpos, patas y alas de insectos, sino una miríada de otros biomorfos, entre ellos ranas arborescentes, murciélagos y flores. Incluso aparecen formas de objetos artificiales, como lámparas y balanzas de precisión [*véase la figura 4*].

¿De qué modo produce RELOJERO esas variaciones sobre un mismo tema? Como he mencionado, la forma de los árboles que genera el programa está controlada por genes. Se dispone, en total, de 16 genes. Los efectos de algunos genes se describen fácilmente. Los de otros, en cambio, resultan impredecibles, por sus interacciones con otros genes. Así, ciertos genes controlan el número de ramas y el tamaño global del árbol. Entre los del segundo

tipo, tres genes se combinan para controlar la extensión horizontal de las ramas, mientras que otros cinco determinan, conjuntamente, su extensión vertical.

Dawkins considera un espacio de 16 dimensiones habitado por todas las variaciones de biomorfos concebibles. De hecho, cada criatura está determinada completamente por un juego de 16 números, que constituyen su dotación genética. Tales números pueden interpretarse como coordenadas, de tal modo que cada biomorfo pasa a ser meramente un punto del que Dawkins denomina país de los biomorfos. RELOJERO es vehículo excelente para la exploración de ese territorio.

El programa de Dawkins es demasiado largo y complicado para explicarlo aquí. En primer lugar, dispone de gran número de opciones, cada una de las cuales puede considerarse, por derecho propio, un programa. Por ejemplo, RELOJERO tiene una opción de ingeniería; los usuarios pueden consultar en el manual explicativo del programa la forma de manipular directamente los genes al objeto de obtener variedades de biomorfos. Entre otras de las opciones con que cuenta está la de un "registro fósil", merced al cual puede almacenar líneas de descendencia, remedando las formaciones de estratos geológicos, y la opción "monstruo optimista", que selecciona al azar la sucesiva progenie. Tal vez aparezca algo interesante; quién sabe si un monstruo.

A estas alturas, el creacionista habrá gritado "¡Juego sucio!", y con derecho, qué duda cabe. ¿Qué hay de deportivo en la selección no natural de la variedad superviviente que va efectuando un humano? En realidad, tal proceder no es más que cría selectiva, pura y simplemente. Darwin se sirvió del éxito de los criadores humanos para poner de manifiesto la variedad de las poblaciones. Pero ello no demostraba la teoría de selección natural; el razonamiento de Darwin, muchísimo más amplio, se fundaba en numerosas fuentes, entre ellas, en las pruebas geológicas. También salta a la vista, por otra parte, que Dawkins ha probado su tesis, a saber: que el efecto de los cambios acumulativos puede ser impresionante. ¡Cuánto más impresionante no sería si pudiera refutarse directamente la objeción de que se ha hecho trampa! En lugar de la selección efectuada por humanos, o incluso mediante procesos de azar, ¿sería posible que el procedimiento de selección de RELOJERO quedara bajo control directo de las presiones ambientales? ¿Podría derivar la morfología de la interacción competitiva entre las especies?

Dawkins ha lanzado a los programadores el siguiente desafío: "Los biomorfos deberían interactuar, en el ordenador, con la simulación de un ambiente hostil. Ciertas peculiaridades de su forma deberían determinar su supervivencia o aniquilación en tal entorno". Dawkins prosigue diciendo que la vulnerabilidad de los organismos debería emerger de la propia simulación, y no ser directamente introducida por el programador. Incluso podría variar el número de genes. Dawkins opina que, de haber alguien a la altura de la tarea, seguramente sea alguno de los "programadores que preparan esos vulgares y ruidosos vídeojuegos de los salones recreativos". No sé por qué, pero dudo de que así sea.

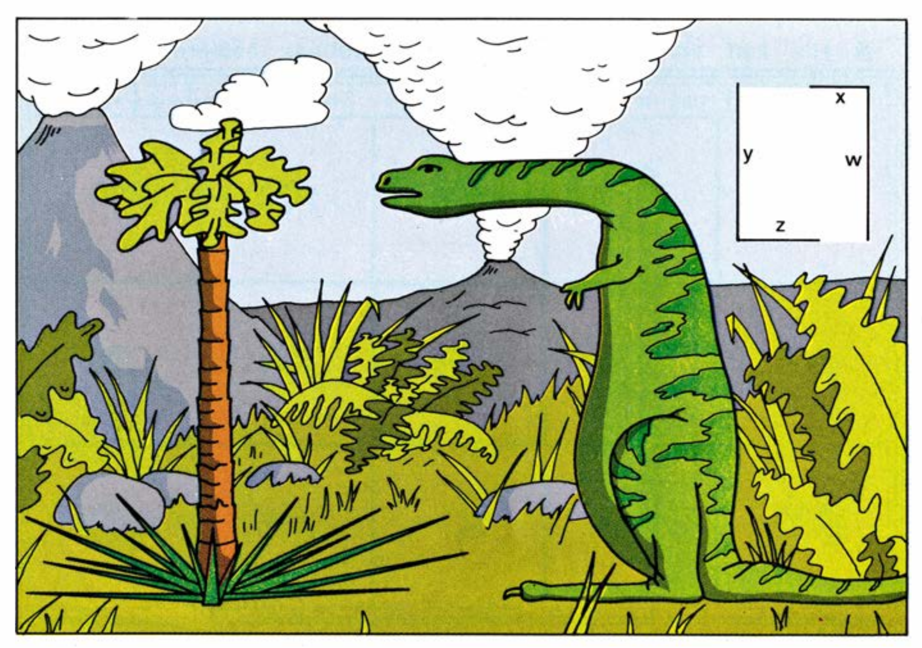
No puedo afrontar personalmente el reto de Dawkins, pero sí ilustrar la clase de situaciones que tiene en mientes. Después de todo, sí es posible diseñar *ab initio* dos especies que interactúen y que cambien de forma a raíz de su interacción. Pero he sido yo, el RELOJERO recreativo, quien ha introducido la vulnerabilidad. El ejemplo siguiente quizás anime a alguien a convertir en un gran salto adelante el pequeño avance que aquí expongo, y transformarlo en un verdadero programa de evolución (por no decir de revolución).

Fijémonos en el ortosaurio y en la espinofita de la figura 1. El ortosaurio no puede sobrevivir sin alimentarse del penacho de exquisitas hojas que coronan la espinofita. Desdichamente para él, el pie de la espinofita está rodeado por un collar de mortales púas. Sin embargo, como su nombre indica, el ortosaurio tiene el cuello doblado en ángulo recto. Peculiaridad anatómica que permite al ortosaurio evitar los mortales pinchos de la espinofita y alcanzar su alimento.

Hasta aquí las buenas nuevas. Las malas son que el cuello del ortosaurio es de una rigidez pavorosa. El ángulo formado por el cuello y el cuerpo es permanentemente recto. Además, tanto el cuello como el cuerpo son rígidos. Así, el ortosaurio se ve obligado a alimentarse de las hojas situadas a cierta altura, y ello, suponiendo que la longitud de su pescuezo sea suficiente para alcanzarlas.

Por supuesto, las poblaciones de ambas criaturas están sujetas a variaciones aleatorias de su dotación genética. Puede variar la altura del ortosaurio (medida hasta la cruz), así como la longitud de su pescuezo. Las puntiagudas hojas de la espinofita tienen todas la misma longitud en cada planta, pero tal longitud puede variar entre plantas. También la altura de la espinofita está sujeta a cambios pequeños y aleatorios.

Bastan dos genes para controlar la forma del biomorfo. Un gen del ortosaurio determina su altura y otro fija la longitud del cuello. La espinofita tiene un gen para la altura y otro para la longitud de las espinas. Bajo ciertas condiciones integradas en el programa, que he llamado ESCAPE, estas dos formas de vida coexisten y evolucionan conjuntamente. Se dispone así de una de las piezas de un posible reloj, ya que no de su constructor. En un sentido genético, a largo plazo, la espinofita trata de escapar del ortosaurio evolucionan-



1. La espinofita y el ortosaurio evolucionan conjuntamente

do en la dirección que dificulte el ser devorada.

Por turnos, ESCAPE va dando ocasión a los biomorfos para que cambien de forma. Por ejemplo, la espinofita puede hacerse más alta o más baja, y sus hojas espinosas pueden alargarse o acortarse. También pueden variar el cuello y el cuerpo del ortosaurio, alargándose o encogiéndose. Se acota por arriba el tamaño de ambos biomorfos, para impedir una carrera de armamentos sin fin. Sucede, por ejemplo, que la suma de la longitud de las espinas más la altura del tronco de la espinofita no puede rebasar cierto límite que he llamado *espinomax*. Análogamente, la longitud del ortosaurio, desde el morro hasta el talón, no puede exceder de la cota *ortomax*. Al objeto de hacer la situación más interesante, opté por hacer que *ortomax* fuese mayor que *espinomax*.

¿Cómo dirime ESCAPE qué descendientes de los biomorfos sobrevivirán? La respuesta reside en una magnitud llamada "distancia trófica", que mide lo lejos que se halla el hocico del ortosaurio de las fragantes hojas de la espinofita. Supongamos que el ortosaurio tenga una altura w y un cuello de longitud x. Sea y la altura de la espinofita y z la longitud de sus espinas. En tal caso, la distancia trófica está dada, sencillamente, por el teorema de Pitágoras; es decir, la raíz cuadrada de la cantidad $(z - x)^2 + (y - w)^2$. Si se diera el caso de ser z menor que x, el ortosaurio podría permanecer tan cerca como quisiera de la espinofita y sólo se tomaría en consideración el segundo de los términos anteriores.

No sorprende que la espinofita tienda a evolucionar en el sentido de hacer máxima la distancia de trofismo. Y a la inversa; el ortosaurio busca hacerla mínima. Llegado su turno, cada biomorfo engendra dos variedades. Por ejemplo, la altura de la espinofita, que en el momento actual es y, puede convertirse en $y - 1$ o en $y + 1$ en la generación siguiente. La longitud de las espinas puede incrementarse de z a $z + 1$, o decrecer a $z - 1$. En cada caso se elige al azar una de las dos variantes. El ortosaurio origina otras dos variantes, exactamente del mismo modo. Las especies paternas se conservan junto a las variantes, para someterlas a evaluación frente a su oponente biológico; la distancia de trofismo proporciona el criterio de valoración. Por ejemplo, entre las tres espinofitas, este proceso cuasinatural seleccionará la que tenga mayor distancia trófica respecto del actual modelo de ortosaurio, y será ésa la que pase al siguiente turno de evolución. Es ahora el turno del ortosaurio. Se ensayan frente a la nueva espinofita sus dos variantes y la estirpe paterna; la que tenga mínima distancia de trofismo se convertirá en nueva y ufana progenitora de su raza.

Sería maravilloso dotar a ESCAPE de una presentación gráfica similar a la de la reconstrucción de los dos antiguos y fabulosos biomorfos realizada por nuestro dibujante. Pero ello rebasa mis habilidades. He optado, en cambio, por un diagrama del tipo más sencillo, como el que acompaña a la figura 1. Apenas se representan en él más que las variables que interesan al caso, plasmadas en líneas rectas. Los lectores que no deseen probar siquiera esa sencilla presentación tendrán que contentarse con hacer que ESCAPE imprima los valores de las cuatro variables al cabo de cada ronda.

ESCAPE se vale de un único bucle, en el seno del cual ambos biomorfos se modifican, evalúan y presentan:

entrar *ortomax*, *espinomax*
entrar w, x, y, z
entrar *gen*
para $i \leftarrow 1$ hasta *gen*
    engendrar variedades de espinofitas
    evaluar las espinofitas
    engendrar variedades de ortosaurios
    evaluar los ortosaurios
    mostrar los ganadores

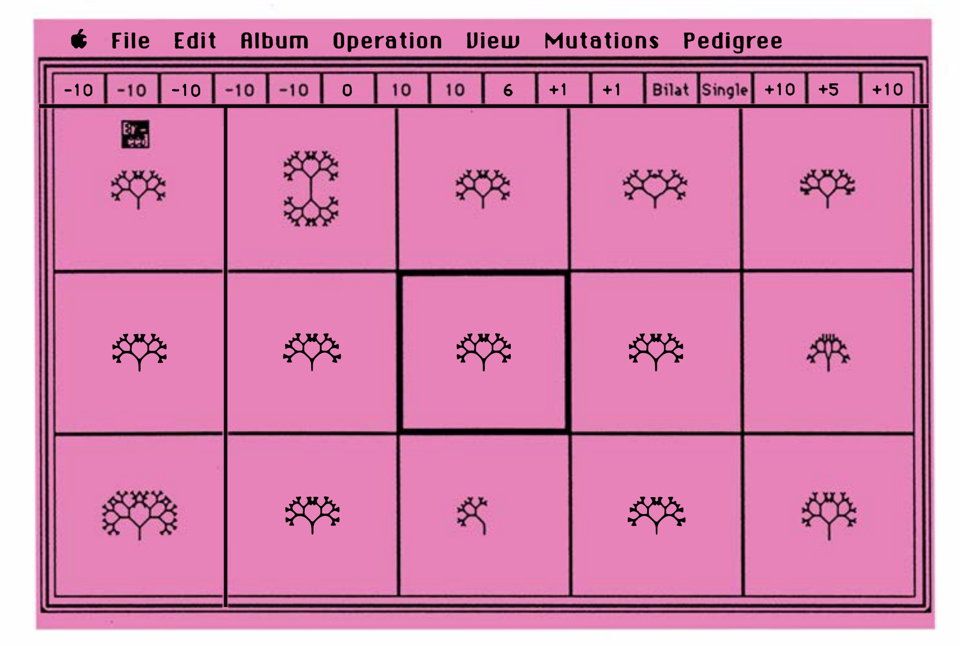
Por supuesto, el usuario ha de entrar a mano los valores de las seis variables clave e introducir igualmente el valor de gen, el número de generaciones, o rondas, que se desea. El bucle procede entonces a iterar repetidamente dos subciclos de reproducción, evaluación y presentación. La reproducción de un biomorfo se produce seleccionando al azar un par de números y utilizándolos para determinar si alguno de los genes se reforzará o debilitará, por así decirlo. Por ejemplo, el código algorítmico que engendra variedades de ortosaurios produce las mutaciones a y b de los genes w y x por la siguiente técnica:

$r \leftarrow$ al azar
si $r < 0{,}5$ entonces $a \leftarrow w - 1$
    en otro caso $a \leftarrow w+1$
    $s \leftarrow$ al azar
    si $s < 0{,}5$ entonces $b \leftarrow x - 1$
        en otro caso $b \leftarrow x + 1$

El programa tiene ahora tres ortosaurios entre manos. Sus dotaciones genéticas pueden representarse mediante tres pares de números: $[w,x]$, $[a,x]$ y $[w,b]$. ¿Cuál es la mejor de las tres? ESCAPE evalúa los modelos calculando sus respectivas distancias de trofismo a la espinofita actual. Si denotamos las tres distancias $d1$, $d2$ y $d3$, podemos servirnos básicamente de una misma fórmula para todas. He aquí, por ejemplo, cómo calcula ESCAPE la segunda distancia, $d2$, entre el ortosaurio $[a,x]$ y la espinofita $[y,z]$:

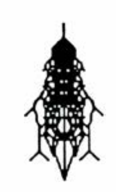
si $x < z$ entonces $d2 \leftarrow (z - x)^2 + (y - a)^2$
    en otro caso $d2 \leftarrow (y - a)^2$

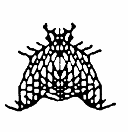
En todos los casos se utiliza la fórmula de la distancia euclídea, con las



2. Variantes genéticas del árbol central







3. Biomorfos zoológicos generados por RELOJERO

coordenadas apropiadas para la boca y las hojas. Una vez calculadas las tres magnitudes, se procede a seleccionar la menor de ellas mediante una serie de comparaciones:

si $d2 \leq d1$ y $d2 \leq d3$
entonces $w \leftarrow a$
si $d3 \leq d1$ y $d3 \leq d2$
entonces $x \leftarrow b$

De no verificarse ninguno de esos dos pares de desigualdades, el progenitor sigue siendo superior a la descendencia. Los genes w y x permanecerán invariables. ESCAPE evalúa las espinofitas del mismo modo, salvo que en tal caso se preferirán las distancias mínimas en vez de las máximas.

En la explicación precedente nada se dice de los límites de crecimiento a los que anteriormente aludí. ¿Como evitará ESCAPE que ningún biomorfo rebase sus límites intrínsecos de tamaño? La forma más sencilla consiste en penalizar al biomorfo en el momento de calcular la función de distancia. Por ejemplo, si el ortosaurio variante $[a,x]$ fuera demasiado grande, la suma $a + x$ superaría la cantidad *ortomax*. En tal caso podemos sumar 1000 a la distancia, para impedir que esa variedad gane en la evaluación. Recíprocamente, para penalizar a las espinofitas se restan 1000. Es preciso penalizar análogamente a ambas criaturas en el caso de que una cualquiera de las cuatro variables se haga negativa.

Los vencedores de los procesos de evaluación se presentan en la pantalla mediante las sencillas órdenes de trazado de rectas de que disponen prácticamente todos los lenguajes de programación. Esa parte del programa queda encomendada a la imaginación del lector.

¿Qué sucede cuando se ejecuta el programa ESCAPE? Ello depende, hasta cierto punto, de los valores iniciales de los cuatro genes y de los valores límites que se le proporcionen al programa. Las variables pueden reflejar igualmente el tamaño de la pantalla. Tomemos como *espinomax* el valor 100 y, como *ortomax*, 110. Esos datos iniciales crean un interesante fenómeno. En principio, demos a los cuatro genes cualquier cifra entre 0 y 50. Podría empezarse igualmente con biomorfos razonablemente bien emparejados, ya espigados y estrechos, ya achaparrados y anchos, por así decirlo. Con la sucesión de las generaciones se produce una carrera. El ortosaurio quizá "persiga" a la espinofita mientras ésta se adentra en la gama de variedades altas, atrapándola por fin en la cumbre. A lo que la espinofita respondería tratando de huir hacia abajo y proyectando espinas cada vez más temibles. El ortosaurio se achaparra en consecuencia, al tiempo que su cuello se desarrolla y alarga. La sencilla condición de que *ortomax* exceda de *espinomax* garantiza un ciclo aparentemente interminable de huidas y persecuciones. Los lectores provistos de imaginación podrán visualizar el resultado cuando se dé el caso contrario.

Como es obvio, las huidas y persecuciones que acabamos de mencionar no se refieren a los diversos ejemplares de biomorfos, sino al registro fósil. Los paleontólogos que examinasen una serie completa de estratos rica en espinofitas y ortosaurios fósiles descubrirían, perplejos, una variación conjunta del tamaño a lo largo del tiempo, a modo de oscilación pendular.

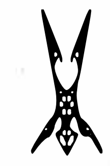
A Dawkins le gustaría tener noticias de los lectores que desarrollen programas experimentales y operativos capaces de explorar el tema de la emergencia de la evolución. Tales programas podrían ser elaboraciones de ESCAPE, o algo mucho más profundo. La dirección de Dawkins es: University of Oxford, Department of Zoology, South Parks Road, Oxford OX1 3PS, Inglaterra.

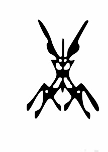
Aunque persiste el debate entre creacionistas y evolucionistas, nos preguntamos hasta qué punto muchos de los argumentos que presentan ambos bandos se basan en prejuicios. Basta con imaginar un Creador de inteligencia suma, perfectamente capaz de crear por evolución, que observa a los contendientes reduciendo a sus adversarios a monigotes.

El juego llamado MAD (destrucción mutua asegurada) fue tema central del artículo de diciembre del año pasado. Desarrollado, en 1985, por Tad Homer-Dixon y Kevin Oliveau en el Instituto de Tecnología de Massachusetts, simula la interacción entre dos superpotencias un tanto indecisas entre la guerra y la paz. En experimentos realizados en el MIT, los estudiantes ocuparon las consolas de mando de sus terminales informáticas y represetaron el papel de jefes de estado. ¿Debería cooperarse, no ir a la guerra y conformarse con un modesto incremento de la puntuación total propia, o mejor ser los primeros en atacar y obtener así —es de presumir— una buena cosecha de puntos? Conforme avanza el juego, las recompensas potenciales

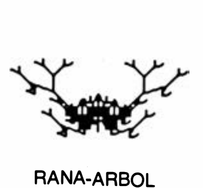




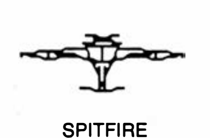














4. Biomorfos y otras formas generadas por RELOJERO

que ofrece la defección (atacar el primero) crecen sin cesar.

Se recomendó a los estudiantes que procuraran hacer máxima su puntuación total (objetivo que favorece la cooperación), pero muchos no se atuvieron a ese espíritu. Ese fue también el caso de David I. Farmer, de Mississauga, Ontario, quien trató de analizar un juego similar aplicando los principios de teoría de juegos, que se ocupa únicamente de matrices de juego consideradas aisladamente, sin tomar en cuenta las jugadas posteriores. La teoría de juegos busca igualmente hacer máxima la puntuación total, para lo cual pretende hacer mínimas las pérdidas en cada turno. Al objeto de aplicar la teoría, Farmer buscó y halló un método para generar nuevas matrices de juego que sean reflejo de las jugadas ulteriores.

Farmer remitió un ejemplo donde intervenían las cinco matrices centrales de MAD. A juicio del lector canadiense, tiene sentido ir a la guerra ya en la segunda matriz. Pero concluye: "Al tener que afrontar la amenaza cada vez mayor de un primer y devastador ataque, para mantener la paz se requiere una lógica que trasciende la teoría de juegos. Sin embargo, si el valor negativo de futuros supuestos bélicos pudiera hacerse menos negativo, incluso la teoría de juegos permitiría evitar la opción bélica. Las conversaciones de desarme contribuyen a reducir el riesgo de una guerra futura y modifican los pagos de la matriz en sentido positivo".

En sus últimas fases, MAD está dominado por una situación conocida por "dilema del prisionero". En este caso, los pagos que reporta la defección son tan grandes que parece lógico pulsar el botón. A pesar de ello, un campeonato de dilema del prisionero celebrado hace algunos años en la Universidad de Michigan reveló que la mejor estrategia para dicho juego era la del "ojo por ojo": en tanto nuestro oponente coopere, hacerlo nosotros también. Los estudiantes (y demás) pueden ya realizar sus propios torneos del "dilema del prisionero", pues acaban de aparecer los programas pertinentes para el Macintosh y pronto lo harán los del IBM PC y compatibles, preparados por Leslie Burkholder y Chris Walton, de la Universidad Carnegie-Mellon. Los lectores interesados pueden indagar cuál es el estado de la cuestión escribiendo a los autores al Center for Design of Educational Computing, Carnegie-Mellon University, 5000 Forbes Avenue, Pittsburgh, Pa. 15213.

Taller y laboratorio

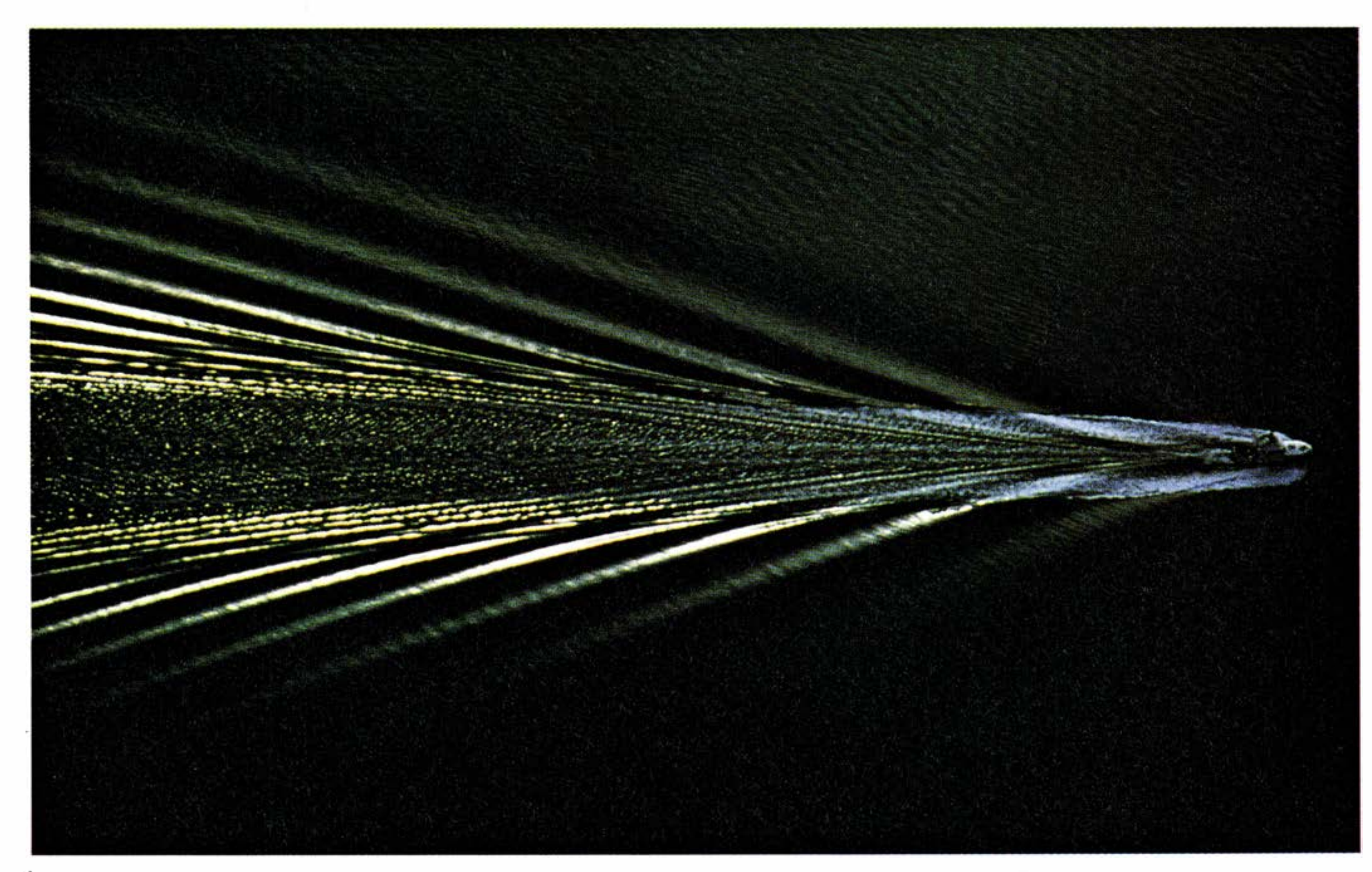
La estela que deja una embarcación es una compleja figura de interferencia

Jearl Walker

Cualquiera que haya volado a relativa altura por encima de una vía navegable se habrá fijado en las formas de las estelas que dejan las embarcaciones en su avance: Dibujan unas figuras intrincadas, constituidas por múltiples olas encerradas en una suerte de V que parece ir a remolque de la embarcación. El aspecto general de la estela podría sugerir que se trata de una onda de choque, similar a las que generan los aviones supersónicos. (Aunque las ondas de choque en el aire son normalmente invisibles, pueden fotografiarse con dispositivos especiales en un túnel aerodinámico.) La onda de choque también tiene forma de V y su vértice parece estar unido al aparato que la crea.

Pese a la semejanza de ambos tipos de figuras, cada una obedece a causas diferentes. Un indicio revelador lo tenemos en el ángulo que forman los brazos de la V. En la onda de choque, este ángulo depende de la velocidad relativa del avión y del sonido a la altitud de vuelo. Si la velocidad del avión aumenta, el ángulo disminuye y la V se estrecha. Conforme la velocidad disminuye, la V se ensancha hasta desaparecer cuando la velocidad del aparato se hace menor que la del sonido.

Si la figura de una estela fuera sólo una onda de choque, el ángulo de la V dependería también de la velocidad. Pero esto no es así: ese ángulo se encuentra siempre en torno a los 39 grados, cualquiera que sea la velocidad de la embarcación. De hecho, absolutamente todos los cuerpos, desde los patos hasta los superpetroleros, dejan al moverse por el agua una estela con ese mismo ángulo. Además, las estelas difieren de las ondas de choque en la complejidad de su estructura. Así, los brazos de la V no son dos ondas únicas y largas, sino que están constituidos por unas series de ondas de corta longitud que son la causa de su aspecto de pluma [*véase figura 2*]. Entre los brazos se extienden ondas transversales curvas



1. Vista aérea de la estela de una motora

que parecen moverse acompañando a la embarcación.

El primero en resolver el mecanismo que genera las estelas de las embarcaciones fue Lord Kelvin, en 1887. Encontrar la solución fue tarea ardua y requirió la invención de un artificio matemático llamado método de fase estacionaria. Este método permitió a Kelvin llegar a una aproximación sobre la manera en que las ondas creadas por una embarcación interfieren entre sí. Las investigaciones acerca de la generación de estelas ha proseguido hasta nuestros días, estimuladas por el hecho de que gran parte de la energía que consumen los barcos se desperdicia en la creación de las estelas.

Cuando nos imaginamos una onda acuática lo más probable es que pensemos en una onda sinusoidal que se propague por la superficie del agua a cierta velocidad. Esta onda de fase, así se la conoce, se caracteriza por una sucesión de crestas y valles uniformemente espaciados, de altura y profundidad constantes. La distancia horizontal entre dos crestas sucesivas es la longitud de onda. La frecuencia con que las crestas sucesivas pasan por un punto fijo constituye la frecuencia de la onda, que es inversamente proporcional a la longitud de onda. La amplitud de la onda es la altura de la cresta respecto a la superficie en calma.

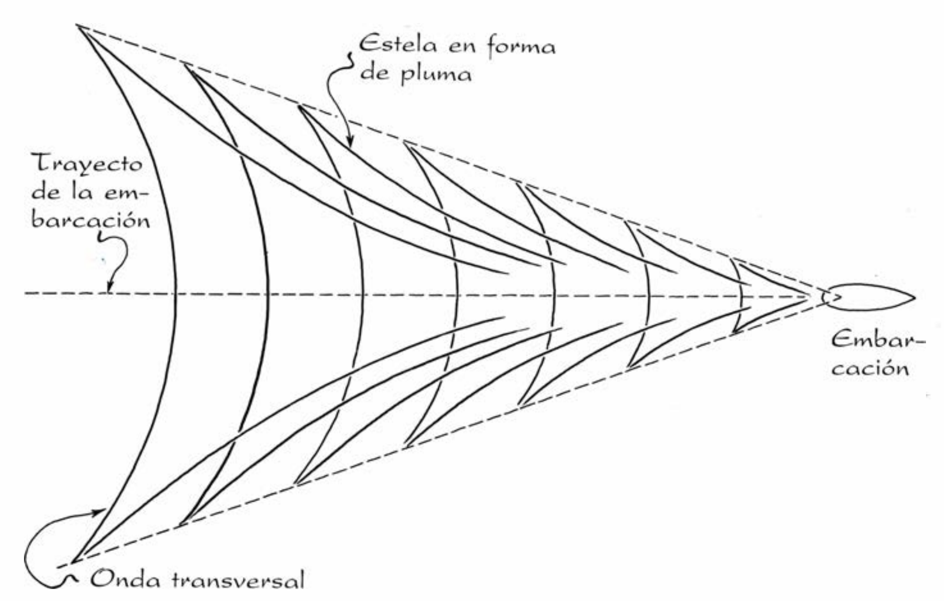
Aunque la onda de fase constituya un modelo sencillo, en realidad nunca se ha visto ninguna en el agua, debido a una sutil propiedad de las ondas acuáticas: la velocidad de éstas depende de su longitud de onda, o lo que es lo mismo, de su frecuencia. Esto explica por qué cuando perturbamos una superficie acuática vemos una figura complicada, y no una onda sinusoidal, viajando por el agua. Esta figura resulta de la interferencia de numerosas ondas de fase, cada una de las cuales se propaga a una velocidad diferente; y aunque estén presentes distintas ondas de fase, quedan ocultas a nuestra observación.

Una onda acuática que sea aproximadamente una onda de fase puede producirse preparando una máquina de modo que fuerce oscilaciones periódicas en el agua a una cierta frecuencia. (También habrá que eliminar la posibilidad de que retornen ondas reflejadas a la zona que vayamos a examinar.) Si nos fijamos entonces en algún punto no demasiado próximo a la máquina, veremos que ésta genera una onda sinusoidal cuya frecuencia es la de la máquina.

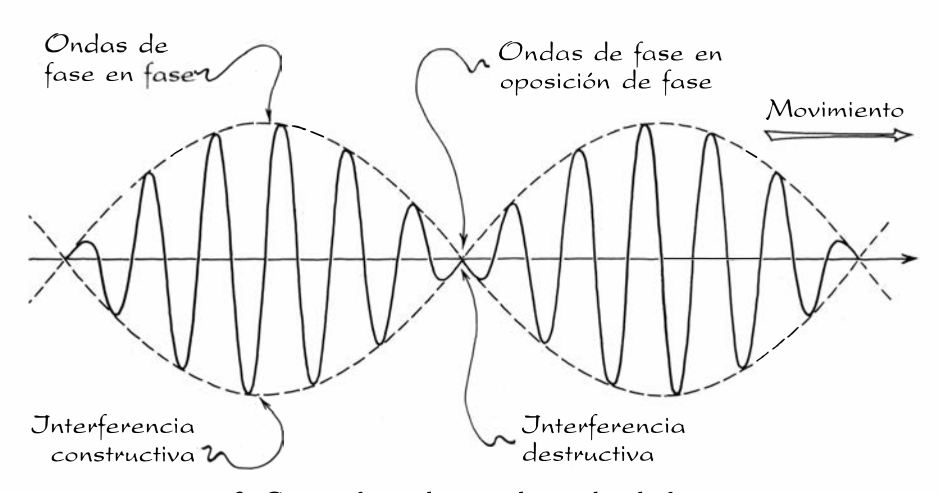
Supongamos que medimos la velocidad de la onda y ajustamos luego la máquina para que genere una onda de mayor frecuencia y, por tanto, de menor longitud de onda. Descubriremos que esta nueva onda se propaga más lentamente que la anterior. Con una experimentación cuidadosa se descubre que la velocidad de una onda de fase depende de la raíz cuadrada de su longitud de onda. Si la velocidad de una onda depende, de un modo u otro, de su longitud de onda, se dice que la onda es dispersiva. Esta propiedad dispersiva del agua confiere a la estela que deja una embarcación su forma característica.

En los libros de texto acostumbra a presentarse el fenómeno de la dispersión con un ejemplo relativo a dos ondas de fase que difieren muy poco en longitud de onda y velocidad. Si estas ondas de fase se propagan por la superficie del agua, interfieren entre sí dando lugar a una figura móvil en la que se esconden las ondas de fase. La superficie del agua adquiere así una estructura ondulada, en la que las ondas se reúnen en grupos que parecen marchar en procesión en la misma dirección que las ondas de fase [*figura 3*]. Crestas y valles avanzan, dentro de cada grupo, con una velocidad doble de la del grupo; aparecen primero en la parte posterior de éste y aumentan de tamaño al acercarse al centro del mismo, para acabar desvaneciéndose al llegar a la parte delantera.

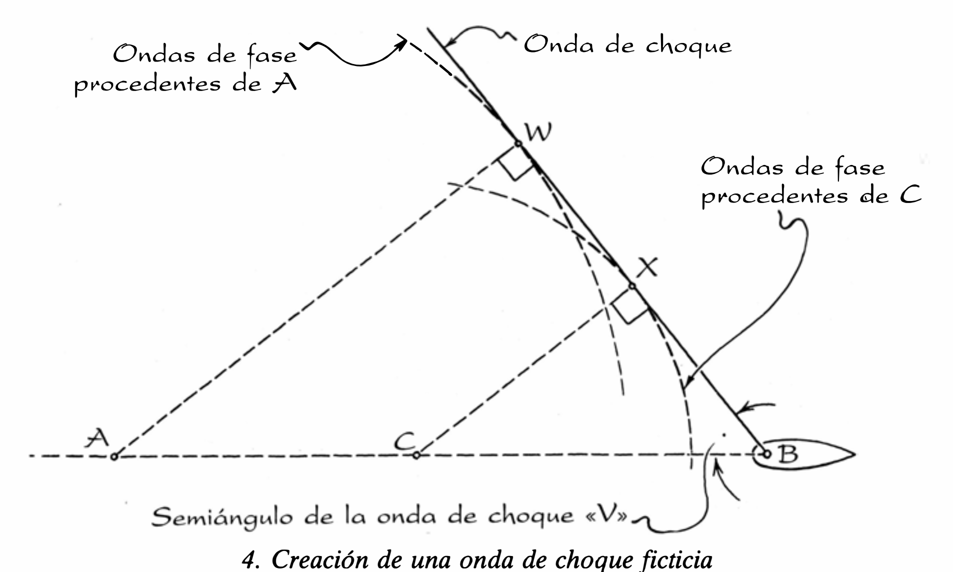
Las crestas y valles señalan los puntos donde coincide que las ondas de fase están aproximadamente en fase, o en sincronía, por lo que se suman constructivamente para formar crestas elevadas y valles profundos. Hacia cada extremo del grupo, las ondas de fase se encuentran parcialmente en posición de fase y se interfieren destructivamente, dando lugar a crestas y valles menos pronunciados. Entre un grupo y otro la



2. Ondas que configuran la estela de una embarcación



3. Grupos formados por dos ondas de fase



4. Creación de una onda de choque ficticia

interferencia destructiva es completa. Téngase en cuenta que, aunque los grupos parecen moverse conjuntamente (en particular si nos olvidamos por un momento de su estructura interna), en realidad se están recreando continuamente conforme nuevos tramos de las ondas de fase pasan por ellos. Lo que vemos moverse y lo que parece persistir son las zonas de interferencia constructiva.

Si las dos ondas de fase se sustituyen por otras cuyas longitudes de onda varíen entre márgenes estrechos, la figura de interferencia es similar, con unas crestas y valles que atraviesan los grupos a una velocidad doble de la de éstos. Sin embargo, esa figura no tarda en diferir de la del ejemplo precedente, porque las ondas de fase de mayor longitud de onda, debido a la dispersión, adelantan a las de longitud de onda más pequeña. Conforme los grupos se alejan del foco de ondas, cambian de forma continuamente. Una manera de definir los grupos consiste en designarlos como aquellas zonas donde las ondas de fase que los componen se hallan siempre en fase; y de ahí la expresión "fase constante" de Kelvin.

En la mayoría de los casos, las perturbaciones de una superficie acuática son más complicadas que en los ejemplos precedentes, porque aparecen ondas de fase cuyas longitudes de onda varían entre márgenes muy amplios. Supongamos que caiga una piedra en un estanque de gran superficie que estamos contemplando desde cierta distancia. Durante un corto tiempo, quizá decenas de segundo, la perturbación emite continuamente ondas de fase. Las de longitud de onda larga (y por tanto las de mayor velocidad) serán las primeras que comenzarán a pasar ante nosotros. Con el tiempo, empezarán a pasar ante nosotros ondas de fase cada vez más lenta. Aunque en poco tiempo pasan una multitud de ondas, la superficie del agua permanecerá en calma, porque todas esas ondas se interfieren destructivamente: cada cresta de una onda de fase dada es anulada por los valles de otras ondas de fase.

Los primeros signos de actividad nos llegan cuando comienzan a pasar los grupos. Cada uno de éstos viaja a una velocidad dos veces menor que sus ondas de fase asociadas. El primer grupo, que es el más rápido, lo producen las ondas de fase cuyas longitudes de onda son las mayores, y por tanto sus velocidades son las más elevadas. Con el tiempo, irán pasando los grupos que viajan más lentamente, asociados a ondas de fase de longitudes de onda y velocidades progresivamente menores. Mientras que en el primer caso se organizaba un desfile de grupos idénticos, la piedra origina una sucesión de grupos que se funden unos con otros.

La transición de un grupo a otro es tan suave que quizá tengamos dificultades en percibirla, ya que además la acción transcurre con rapidez ante nuestra mirada. Si fotografiamos el despliegue de ondulaciones con el propósito de inmovilizar la acción y medir las separaciones entre cresta y cresta, veremos que los grupos más veloces van por delante de los lentos. La separación es mayor en la zona más alejada del punto de impacto de la piedra, lo que indica que las ondas de fase con mayores longitudes de onda son las responsables de esa zona. La separación se va acortando conforme examinamos el agua y nos acercamos al punto del impacto, lo que indica que en esa zona entran en juego las ondas de fase con longitudes de onda menores. Haciendo una serie de fotografías podremos comprobar que los grupos se dilatan a medida que las ondas de fase rápidas toman la delantera a las más lentas, hasta que finalmente las pérdidas energéticas y la propagación de las ondas eliminan toda acción.

Una piedra produce una perturbación en un punto único. Una embarcación en movimiento da lugar a una serie continua de perturbaciones puntuales, cada una de las cuales emite grupos que se propagan e interfieren entre sí formando la estela característica. Solucionar el problema de la interferencia tal como hizo Kelvin es asunto complejo, por lo que comenzaré con una exposición mucho más sencilla que, en 1984, presentó Frank S. Crawford, de la Universidad de California en Berkeley.

Crawford empieza por considerar la onda de choque que produciría una embarcación en movimiento si las ondas acuáticas no fueran dispersivas. En la figura 4 tenemos una embarcación que se ha desplazado desde *A* hasta *B* a velocidad constante. La perturbación que creó en *A* emitió unas ondas de fase circulares (moviéndose todas a la misma velocidad, ya que pasamos por alto la dispersión). La embarcación creó asimismo ondas de fase en todos los puntos entre *A* y *B*, generando una figura que es la combinación resultante de los circuitos de tamaños variables cuyos límites definen una onda de choque en forma de V. Por convenio, el ángulo abarcado por una onda de choque se mide dando el semiángulo de la V. En nuestra figura el seno de ese ángulo es igual al cociente entre la velocidad de las ondas y la velocidad de la embarcación.

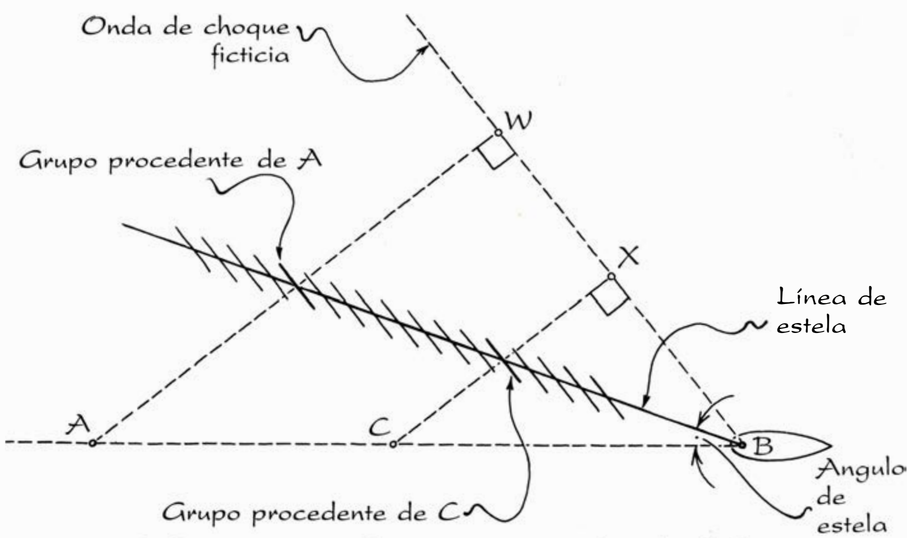
Esta onda de choque ficticia constituirá nuestro punto de partida en la descripción de la estela real a que da lugar la dispersión. Aunque la perturbación producida por la embarcación genera ondas de fase dentro de unos márgenes muy amplios de longitudes de onda, evidentemente las únicas importantes serán aquellas cuyas velocidades no superen la de la embarcación. Consideremos en primer lugar un pequeño paquete de ondas de fase cuyas longitudes de onda sean muy parecidas y que viajen a 0,866 veces la velocidad de la embarcación o valores próximos a ese. Cuando la embarcación llegue a *B*, las ondas de fase generadas en *A* lle-

garan al punto W, situado en el brazo de una onda de choque ficticia que —de acuerdo con la ecuación— forma un ángulo de 60 grados con la trayectoria de la embarcación. Los demás paquetes de esas características que las perturbaciones hubieran producido entre A y B (en el punto C, por ejemplo) llegarán también al brazo de la onda de choque ficticia en ese preciso instante.

Esta onda de choque se llama ficticia porque a lo largo del brazo el agua está completamente en calma, debido a la mutua interferencia destructiva entre las ondas de fase que allí tiene lugar. La actividad que realmente crean las ondas de fase se encuentra por detrás, en los grupos que han viajado hacia fuera a la mitad de velocidad que las ondas de fase [*véase la figura 5*], de manera que el grupo asociado a las ondas de fase que estén llegando a W se encontrará a mitad de camino entre A y W. Trácese una recta que una B con ese punto medio. Esa recta, que podemos llamar línea de estela, señala la posición instantánea de los grupos asociados a las ondas de fase elegidas que se generaron entre los puntos A y B.

Si la embarcación sólo hubiera producido esas ondas de fase que hemos elegido, la estela estaría limitada por la línea de estela, y el ángulo de estela, medido entre esa línea y el trayecto de la embarcación, sería de 19,1 grados. El cálculo del ángulo de estela se efectúa como sigue. Sea L la distancia AB. Como las ondas de fase se eligieron de modo que se propagaran a 0,866 veces la velocidad de la embarcación, W debe estar a 0,866 L de A y el punto equidistante de A y W debe estar a 0,433 L de A. Como el triángulo rectángulo que forman A, W y B tiene un ángulo de 60 grados, el ángulo WAB debe medir 30 grados. El ángulo de estela podrá calcularse directamente aplicando las fórmulas trigonométricas al triángulo creado por A, B y el punto medio de AW.

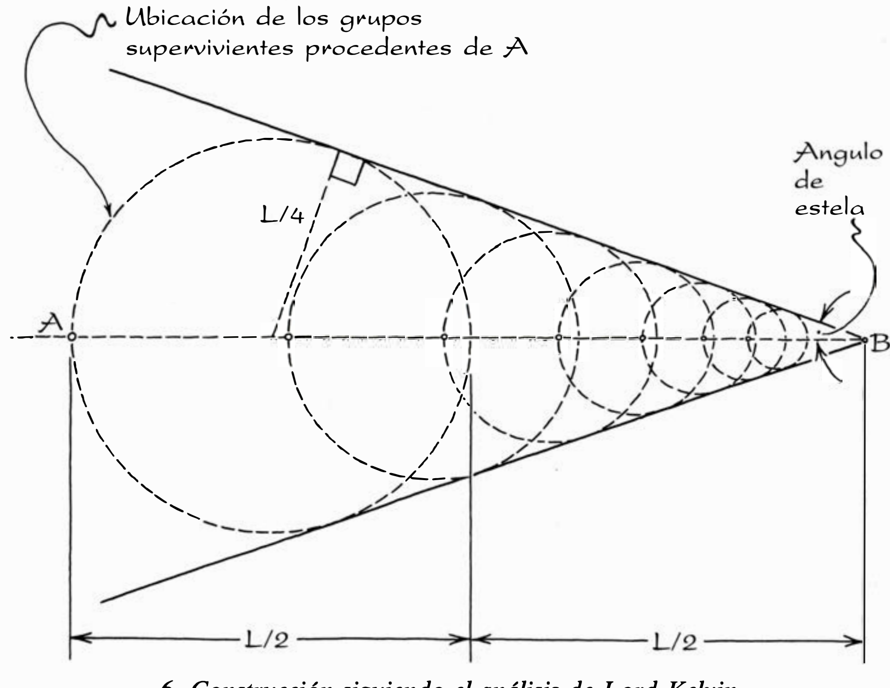
Calculando los ángulos de estela asociados a otras velocidades de onda de fase elegidas, se encuentra que el máximo ángulo de estela tiene un valor de 19,5 grados, que corresponde a las ondas de fase que viajen a 0,82 veces la velocidad de la embarcación. La línea de estela correspondiente a este ángulo marca los límites de la figura. Los límites se distinguen perfectamente, pues se generan unos valles y crestas muy acusados debido a que una amplia gama de ondas de fase contribuye allí a la formación de los grupos. El ángulo total del vértice de la V, doble que el ángulo de estela máximo, es de 39 grados, valor que coincide con el observado en las estelas reales.



5. Grupos correspondientes a un paquete de ondas de fase

Crawford pasó luego a considerar las ondas transversales que aparecen en la figura de la estela. Estas ondas se deben a la propagación de los grupos en arcos circulares en dirección a la embarcación. Como la figura permanece estacionaria con respecto a la embarcación, debe estar dominada por un paquete de ondas de fase que viajen a la velocidad de la embarcación, o un poco por encima o por debajo de la misma. Por supuesto, el grupo asociado viaja a la mitad de velocidad. El paquete creado en A dará lugar a un grupo que estará a medio camino de AB cuando la embarcación llegue a B. Los otros grupos que forman las ondas transversales presentan un comportamiento similar respecto a la embarcación. Si en un momento dado un grupo se encuentra a una distancia x por detrás de la embarcación, el punto de perturbación que lo creó se encontrará entonces a una distancia $2x$ por detrás de la embarcación. Además, la onda transversal que forma ese grupo deberá tener en ese instante un radio x. La figura resultante aparenta ir a rastras de la embarcación, pero lo que realmente pasa es que la renuevan sin cesar otras ondas de fase a medida que se propagan las ondas transversales.

En 1957 James Johnston Stoker, de



6. Construcción siguiendo el análisis de Lord Kelvin

la Universidad de Nueva York, y Douglas Keith, de la Universidad de Bristol, publicaron, cada uno por su cuenta, sendas versiones simplificadas del análisis de Kelvin. En estas versiones los grupos procedentes de *A* interfieren con grupos procedentes de otros puntos. Cuando la embarcación llega a *B*, los únicos grupos supervivientes de los procedentes de *A* se encuentran dentro de una circunferencia que incluye el punto *A* en la parte posterior (con respecto al movimiento de la embarcación). El grupo de la parte delantera del círculo, producido por ondas de fase que viajan a la velocidad de la embarcación, se encuentra a medio camino entre *A* y *B*. Las demás zonas del círculo están dominadas por grupos asociados a ondas de fase más lentas.

Estos círculos no son ondas; se limitan a señalar la ubicación de los grupos. Para determinar la orientación de un grupo en cualquier punto de un círculo, trácese un pequeño segmento rectilíneo que pase por él de modo que sea perpendicular a la recta que une ese punto con el punto de perturbación asociado al círculo. El pequeño segmento rectilíneo representa las ondas visibles del grupo en ese punto del círculo.

En la figura 6 se representan varios círculos, cada uno de los cuales está vinculado a un punto de perturbación situado en su parte posterior. Los límites de la estela quedan definidos por las líneas rectas tangentes a los círculos. Para calcular la inclinación de esas rectas, que es el ángulo de estela, se elige un círculo. El punto de perturbación *A*, a la distancia *L* de la embarcación, se halla en un círculo de radio $L/4$ cuyo centro está a una distancia $3L/4$ de la embarcación. Obsérvese que un radio y la recta que une el centro con la embarcación constituyen los catetos de un triángulo rectángulo. El seno del ángulo de estela, que queda definido por ese triángulo, deberá ser igual al cociente de las longitudes de esas rectas (1/3) de lo que resulta que ese ángulo valdrá 19 grados y 28 minutos.

Hemos considerado hasta aquí las figuras que forman las estelas de las embarcaciones si navegan en línea recta sobre aguas profundas. ¿Cómo varían esas figuras en los bajíos o en aguas de profundidad variable? ¿Por qué cambian, de una figura de estela a la siguiente, la visibilidad de las ondas transversales y las ondas que forman el dibujo? Dejo estos temas al lector para que los estudie la próxima vez que vuele sobre un curso navegable.

Libros

Gardner, energía de la reproducción, termodinámica y teoría de la ciencia

Miguel de Guzmán, Luis Alonso, David Jou y Javier de Lorenzo

ROSQUILLAS ANUDADAS Y OTRAS AMENIDADES MATEMÁTICAS, por Martin Gardner. Editorial Labor; Barcelona, 1987. Esta nueva obra de Martin Gardner invita especialmente por sus mismas características a reflexionar sobre las profundas relaciones entre el juego y la matemática.

La actividad lúdica del hombre es ciertamente compleja. El sociólogo holandés Huizinga ha analizado profundamente en su obra *Homo ludens* el sentido y el origen de las diversas componentes del juego y su misión en el conjunto de la actividad humana. El juego es una actividad libre de intencionalidades utilitarias, un ejercicio que se persigue por sí mismo. Con una función muy importante sin embargo en el desarrollo del hombre, ya que, aunque no se ejecute intencionadamente por tal motivo, constituye una ejercitación instintiva de multitud de potencialidades de diverso tipo, físicas, intelectuales, psicológicas, que preparan al hombre, así como también al animal, para mejor llevar a cabo multitud de tareas que ha de realizar en su mundo real.

Lo lúdico no se opone a lo serio. El juego no es broma, y el peor de los reventajuegos es precisamente aquel que no toma en serio su parte en el juego. La línea que separa el juego del arte es ciertamente confusa y tenue y en todo juego suele haber una cierta componente estética. Como en el arte, el placer derivado del juego se produce a través de su ejecución y a través también de su contemplación.

El juego coloca al jugador en una esfera separada de la vida ordinaria. Como si por el momento, por obra y gracia de su voluntad lúdica, se deshiciese de la sordidez rutinaria de las reglas que rigen la vida común para introducirse en un mundo en que otro reglamento más simple, más fácil, más inteligible, de belleza más asequible, gobernase la existencia.

En el buen juego hay siempre ciertos elementos de tensión, creada por la existencia de adversarios o por la presencia de dificultades de la situación misma en que nos colocan las reglas, tensión cuyo relajamiento conduce precisamente al placer específico que la actividad lúdica proporciona.

El juego da origen a lazos especiales entre los practicantes de la misma actividad que se sienten solidarios por haber participado, separados del resto del mundo en el tiempo y en el espacio de la misma experiencia lúdica. El juego, al fin y al cabo, a través de sus reglas, les ha transportado solidariamente a un nuevo orden, a una nueva vida, llena de ritmo y armonía.

La matemática por su parte es una actividad muy polivalente. Es ciertamente una ciencia, conocimiento sistemático en busca de cierto tipo de leyes que rigen el universo mental y real. Es también un instrumento, una herramienta para la exploración y el dominio del universo, creadora de nuevas técnicas y métodos para utilizar más adecuadamente las fuerzas de la naturaleza. Es también una actividad creadora de belleza, un arte cuya belleza es, como Platón afirmaba, aprehensible "por los ojos del alma". Pero la matemática también participa, y en gran medida, de las características que hemos visto presentes en el juego. La matemática es un gran juego.

Se puede afirmar que los aspectos lúdicos de la matemática son consubstanciales a su misma vida. Muchos campos de la matemática son como la combinatoria, la topología, la probabilidad, los grafos...; han nacido en el juego y con el juego prosperan y se hacen más profundos. Muy frecuentemente el mundo así iluminado de tales juegos resulta coincidir en algunas facetas con aspectos particulares del mundo real y el matemático se alegra entonces de que su exploración lúdica se haya convertido en útil, pero esto no es más que una especie de premio añadido a su placer inicial, libre de intenciones utilitarias. Estas coincidencias vienen a corroborar lo que el matemático en su juego inicial siempre barrunta, aun sin entender muy bien por qué: que en el fondo las reglas autoimpuestas de su mundo lúdico no difieren mucho de las leyes que parecen regir características importantes del mundo real. En esta confianza el matemático puede entregarse a su juego, sabiéndose creador de mundos que, si por el momento parecen tener tan sólo una realidad mental, podrán adquirir corporeidad en cierta medida con el tiempo, cuando la exploración de la realidad sea lo suficientemente profunda para llegar al nivel que él ya ha intuido.

Los juegos de matiz intelectual y la matemática poseen también elementos estructurales muy similares. Las piezas o los elementos de un juego son en matemáticas los objetos iniciales de una teoría. Las piezas se definen a través de reglas, del mismo modo que los objetos matemáticos se definen implícitamente a través de los axiomas. Las buenas jugadas, técnicas, estrategias generales del juego corresponden a los lemas útiles, los buenos teoremas, la teoría elaborada del campo matemático. Los juegos simples con pocos elementos y reglas sencillas tienen su parangón en las teorías matemáticas elementales, en las que se permanece cerca de la base inicial. Los juegos complejos, de complicadas reglas y teorías, son comparables con las teorías matemáticas avanzadas.

La disciplina de aprendizaje y adquisición de técnicas es también muy semejante en las matemáticas y en los juegos. Uno aprende a familiarizarse con las piezas y reglas que son las definiciones y procedimientos iniciales de la teoría matemática. Hay que practicar a continuación las estrategias fundamentales mediante el ejercicio de jugadas simples, que en matemáticas son los ejercicios elementales, las rutinas de la teoría que el tiempo ha consagrado. Para llegar a ser proficiente en el juego, uno ha de observar las jugadas de los grandes jugadores a fin de asimilar en lo posible sus estrategias de pensamiento. En matemáticas hay que familiarizarse profundamente con los grandes teoremas y métodos de la teo-

ría en cuestión, a fin de llegar a incorporar en uno mismo los modos de proceder de los verdaderamente creadores. Finalmente, uno trata de resolver de forma original los problemas abiertos del juego o de la teoría matemática, creando para ello métodos y visiones nuevas.

Debido a esta profunda semejanza entre matemáticas y juegos, no resulta nada extraño observar cómo, a lo largo de la historia, muchos de entre los grandes matemáticos han sido asiduos practicantes de diversas formas de juego. Fermat fue muy aficionado a los cuadrados mágicos, sobre los cuales diseñó toda una teoría, Leibniz fue un entusiasta defensor de la actividad lúdica como fuente de problemas e ideas matemáticas, Euler fue un magnífico jugador de ajedrez, Gauss fue muy aficionado a jugar a las cartas, lo que le servía al tiempo como campo de observación estadística y probabilística, Hamilton inventó y comercializó un juego topológico que se hizo famoso en su tiempo...

En el conjunto de la obra de matemática recreativa de Martin Gardner se puede observar una clara evolución desde una exposición ingenua y sin gran sofisticación matemática, como en sus trabajos iniciales *Mathematics, Magic and Mystery* (Dover, New York, 1956) o *Science Puzzlers* (Scholastic Book Services, New York, 1957), muy en la línea de los puzzlistas clásicos, hasta la publicación de las obras más recientes como la que nos ocupa, en la que muchos de sus capítulos tratan de problemas matemáticos de gran actualidad e interés incluso al nivel de los matemáticos profesionales. Algunos de tales problemas están aún muy lejos de ser resueltos, a pesar de los numerosos esfuerzos recientes.

La estructura de esta undécima colección de artículos es la usual en las más recientes de Gardner. Contiene 21 capítulos, uno de ellos, sobre la paradoja de Newcomb, debido a Robert Nozick. Unos pocos están dedicados a proponer una pequeña colección de problemas recreativos estimulantes y novedosos. Los otros se ocupan de temas diversos. En cada capítulo se mezcla la información técnica con la relativa a su contorno filosófico, histórico, anecdótico, con la habilidad peculiar a la que Gardner tiene acostumbrados a sus lectores. Una sección de complementos y otra bibliográfica en casi cada capítulo redondean de modo muy satisfactorio el tema. La versión en castellano por Labor tiene una presentación muy esmerada, la traducción de Luis Bou es muy fluida y agradablemente legible.

A mi parecer los fragmentos más interesantes y profundos de esta colección son los de naturaleza topológica tales como el capítulo 5 sobre *Rosquillas: En cadeneta y anudadas*, que da su nombre a la obra, el 9, *Sim, Tragón y Pista de Carreras*, en parte basado en el profundo teorema de Ramsey en la teoría de grafos, el 11, *Números de cruces*, acerca de un problema topológico aún por resolver y el 12 sobre *Conjuntos de puntos sobre la esfera*, un tema también en parte pendiente de solución. La pertenencia de tales temas a la matemática viva se evidencia con las referencias continuas al trabajo de muchos matemáticos bien conocidos contemporáneos como P. Erdös, R. Guy, J.G. Mikusinski, J.H. Conway, H.T. Croft...

Obras como la presente de Gardner, juntamente con otras semejantes, como la de Berlekamp, Conway y Guy, titulada *Winning Ways for your Mathematical Plays* (Academic Press, London, 1982) de gran envergadura, parecen ir respondiendo al reto lanzado hace más de dos siglos y medio por Leibniz: "Nunca son los hombres más ingeniosos que en la invención de los juegos... Sería deseable que se hiciese un curso entero de juegos, tratados matemáticamente". Tal vez si nuestra enseñanza de la matemática, a todos los niveles, supiera incorporar con verdadera profundidad el espíritu de los juegos nuestros estudiantes vivirían bajo otra luz su dedicación a la matemática. (M.G.)

REPRODUCTIVE ENERGETICS IN MAMMALS, dirigido por A.S.I. London y P.A. Racey. Clarendon Press; Oxford, 1987. Se recogen aquí las conferencias del simposio sobre la inversión de energía en el proceso reproductor y de amamantamiento de los mamíferos, celebrado en la Sociedad Zoológica Londinense del 10 al 11 de abril de 1986. Se trata de un tema pluridisciplinar en el sentido menos prostituido del término, donde ecólogos y fisiólogos convergen porque a ambos compete averiguar el flujo de energía entre las poblaciones y los costes que en esa moneda hay que asignar a la preñez y lactancia de las crías.

Sabido es que la reproducción de la mayoría de los mamíferos viene regulada por limitaciones nutritivas. Pero apenas conocemos nada de los mecanismos fisiológicos en virtud de los cuales la nutrición regula directamente la fecundidad y separa o acorta los partos y se engendran más o menos crías. También en este dominio la ciencia comienza a andar por pasos; en el simposio de marras, estudiando grupo a grupo la relación entre tasa metabólica basal y energía invertida en la procreación.

Muy oportunamente, precede al grueso del libro un glosario de los principales términos técnicos empleados (tasa metabólica basal, tasa metabólica de campo o de vida libre, tasa metabólica de semicautividad, rendimiento digestivo, rendimiento de asimilación, rendimiento de desarrollo, etcétera) con sus definiciones y unidades respectivas. Unidades que nos indican ya la historia lejana de esta ciencia de sistematización en ciernes. Sin perderse entre iatroquímicos e iatromecánicos, incluso antes, R.M. Laws evoca en una introducción histórica los trabajos de Lavoisier y Laplace sobre combustión química y metabolismo, la aplicación del calorímetro a los animales por Crawford hace ahora doscientos años, los estudios pioneros de Rubner (1883) sobre tasas metabólicas y exigencias nutritivas y térmicas, la relación entre tasas metabólicas y tamaño corporal establecida por Kleiber (1932), la famosa curva "ratón-elefante", con una pendiente de 0,734, elaborada por Brody, Procter y Ashwurth (1934) y las publicaciones clásicas de Benedict (*The Physiology of the Elephant*, 1936), Brody (*Bioenergetics and Growth*, 1945) y Kleiber (*The Fire of Life*, 1961). Esto, por lo que respecta al aspecto fisiológico. En la vertiente propiamente ecológica habría que recordar las investigaciones sobre demografía, redes tróficas y diagramas de flujo a través de los ecosistemas, de cuyo desarrollo histórico nos ocupamos en estas mismas páginas en el mes de mayo de 1987. Y en lo concerniente a la energética química no podemos olvidar, por ejemplo, el hallazgo del ciclo de Krebs (1957). Tales son las grandes corrientes de ciencia básica necesarias para sistematizar la nueva disciplina.

A ellas hay que añadir útiles teóricos propios, como la alometría de los procesos energéticos, y herramientas técnicas: calorimetría indirecta para medir el intercambio gaseoso, técnicas radiactivas e isotópicas para cuantificar los flujos de materia y energía, telemetría para seguir los movimientos de los animales y controlar su consumo energético y sistemas informáticos para las simulaciones.

¿Cómo proceder en esta nueva disciplina? De múltiples maneras. Una, por ejemplo, partiendo de un problema nuclear: la prolongada gestación y breve amamantamiento de los euterios (mamíferos placentarios) se ha considerado un avance evolutivo sobre la corta gestación y largo amamantamiento de los metaterios (marsupiales). ¿Es ello cierto? Expuesto pormenorizadamente el problema, se aborda su solución en clave energética, esto es, en tasas de metabolismo basal. Las conclusiones a las que se llega son muy distintas de las esperadas *a priori* (7-39).

En una primera fase es obvio que predominen los estudios comparativos para poder establecer parámetros estadísticos (curvas de regresión, correlaciones o varianzas) entre distintos conceptos (desarrollo y lactancia, tamaño corporal e inversión de la madre durante la preñez) y entre distintos grupos (entre cánidos, por ejemplo). Comparación que exige conocer primero los términos a cotejar. De ambos grupos de problemas —comparativos y estrategias y pautas de diferentes grupos animales— se trata por extenso en esta obra, que abarca también un detenido estudio sobre la energética de la reproducción humana (275-304). (L.A.)

STATISTICAL THERMODYNAMICS OF NONEQUILIBRIUM PROCESSES, por Joel Keizer Springer Verlag; New York, 1987. La obra que comentamos es probablemente el libro más ambicioso e importante de la década sobre termodinámica de los procesos irreversibles. En ella se consigue una síntesis de las descripciones de Onsager y de Boltzmann relativas a dichos procesos. El punto de conexión viene proporcionado por la consideración de las fluctuaciones.

La teoría de Onsager constituyó la primera teoría completa de fluctuaciones en sistemas en equilibrio. A partir de la teoría de Einstein de la probabilidad de las fluctuaciones, y con una hipótesis de linealidad para las ecuaciones de relajación de dichas fluctuaciones hacia el equilibrio, Onsager consiguió, en 1931, llegar a sus famosas relaciones de reciprocidad entre coeficientes macroscópicos de transporte. Fue la primera utilización de la simetría de inversión temporal para conseguir resultados macroscópicos experimentales. Por este trabajo, que constituye el resultado más importante de la termodinámica de procesos irreversibles clásica, Onsager recibió el premio Nobel de química en 1977.

Una descripción muy diferente de los procesos de transporte fue proporcionada por Boltzmann en 1872, con su famosa ecuación de evolución de la función de distribución de velocidades moleculares. El propósito de Boltzmann era intentar comprender la segunda ley de la termodinámica desde un punto de vista mecánico. Su teoría, basada directamente en las colisiones moleculares, es decir, elaborada a un nivel microscópico, permite evaluar los coeficientes de transporte.

Las diferencias entre las teorías de Onsager y de Boltzmann son muy notables. La primera es más próxima a la termodinámica, incluye fluctuaciones alrededor de los valores medios de las magnitudes y las consideraciones microscópicas están reducidas a la mínima expresión. La segunda es propiamente microscópica, se refiere tan sólo a la evolución media de la función de distribución y permite llegar a valores concretos de los coeficientes de transporte. La teoría de Onsager sólo suministra relaciones formales generales entre ellos, sin permitir su cálculo.

En su obra, el profesor Keizer, de la Universidad de California en Davis, aprehende las características comunes de estas dos descripciones aparentemente tan dispares. Para ello, hace notar que ambas constituyen en realidad contracciones de la descripción puramente mecánica; por lo cual y pese a su carácter microscópico, la descripción de Boltzmann no es mecánica. Le da ello pie a considerar la ecuación de Boltzmann como una ecuación de transporte al estilo de las de Onsager, aunque no en el espacio real, sino en el de las posiciones y momentos moleculares, con la sección eficaz de colisión como único coeficiente de transporte. Una segunda relación entre ambas es la importancia de la entropía, un ingrediente necesario desde el principio en la descripción de Onsager y una consecuencia, en cambio, en la descripción de Boltzmann. El tercer punto es el carácter estocástico, a diferente nivel (fluctuaciones macroscópicas, colisiones moleculares) de ambas teorías.

Sobre estas características comunes, Keizer edifica su teoría, que modifica a la vez sus dos puntos de partida. La ecuación de Boltzmann gana ahora términos fluctuantes, que dan cuenta de las desviaciones de la función de distribución en torno a sus valores medios (idea explorada anteriormente por R.N. Fox, con quien Keizer trabajó en una primera etapa). Por otro lado, se modifica la descripción de Onsager de suerte que sea utilizable fuera de equilibrio; para ello deben añadírsele términos no lineales y librarla de su necesidad *a priori* de la entropía. Ambas modificaciones distan de ser triviales y enriquecen sustancialmente las teorías de partida.

La descripción unificada se lleva a cabo relacionando procesos elementales de transporte (colisiones, reacciones elementales) y fluctuaciones, en una forma canónica en que las variables termodinámicas intensivas (temperatura, presión, potencial químico) influyen en el ritmo de los procesos elementales de variación de las variables termodinámicas extensivas (energía, volumen, composición química). La expresión de esta relación nos recuerda, con ciertas diferencias, la teoría de tasas de reacción de Eyring, que curiosamente no merece mención por parte del autor. Esta relación entre disipación y fluctuaciones es una manifestación más de la vía, enormemente fructífera para la mecánica estadística de fuera de equilibrio, iniciada con los teoremas de fluctuación-disipación de Callen, Kubo y Green en los años 1950.

Se aplica a continuación este esquema a un vasto dominio de situaciones: reacciones químicas, transporte electroquímico en membranas biológicas e hidrodinámica de fluidos simples, etcétera. Finalmente, se presta atención a la termodinámica de estados estacionarios fuera de equilibrio. La diferencia con la descripción clásica basada en la hipótesis de equilibrio local estriba en las nuevas expresiones a que conduce la teoría para los potenciales termodinámicos (entropía, energía libre) y para las ecuaciones de estado fuera de equilibrio.

El libro recoge y sistematiza los resultados que el autor ha ido publicando desde 1975 en diversas revistas, especialmente en *Journal of Chemical Physics*. Más que intentar una visión de conjunto de los progresos efectuados en los diversos problemas que se tratan, el autor se ha ceñido a sus propios desarrollos. La posible pérdida de informaciones interesantes en este aspecto queda compensada sobradamente por la coherencia global de la presentación.

El libro interesará profundamente a los estudiosos de los procesos de transporte, que encontrarán una obra de una ambición fuera de lo común, que aporta no tan sólo novedades de interés en algunas aplicaciones, sino que apunta a la modificación de los propios principios de la teoría. A los veinticinco

años de la publicación del texto ya clásico de S.R. De Groot y P. Mazur sobre termodinámica fuera de equilibrio (North Holland, Amsterdam, 1962, reimpreso recientemente por Dover), el texto de Keizer es la mejor celebración posible en este campo, poco espectacular pero sólido y útil, de la termodinámica de los procesos irreversibles. (D.J.)

Los Nuevos Redentores. Reflexiones sobre la Ingeniería Genética, la Sociobiología y el Mundo Feliz que nos prometen, por José Sanmartín. Ed. Anthropos, en colaboración con la Ed. Universitaria del País Vasco; Barcelona, 1987. La imagen clásica de la ciencia, y de aquel que la cultiva, el científico, va ligada a la visión de una búsqueda desinteresada del conocimiento, de la verdad. La ciencia, que se manifiesta en teoría, es captación, comprensión y explicación de los fenómenos de la naturaleza. En ocasiones, los resultados obtenidos pueden ser aplicados tanto en beneficio como en perjuicio del hombre. Pero esta aplicación, en uno y otro caso, no depende del científico ni de su producto teórico, sino del técnico y de los poderes políticos. Frente al conocimiento desinteresado, la aplicación técnica, la transformación productiva de la naturaleza y el aspecto crematístico y de poder. En este cuadro de visión clásica, la filosofía de la ciencia se ha pretendido un estudio, también teórico y objetivo, de las metodologías y formalismos de las teorías con que los científicos crean y plasman sus trabajos.

Sin embargo, desde hace algunos años, y como consecuencia de efectos como la bomba atómica, desarrollos como las computadoras, aplicaciones como los pesticidas y las hormonas..., se ha venido cuestionando una concepción como la antes apuntada. Las siempre discutidas relaciones "científico teórico-técnico", "teoría-técnica", "ciencia-poder", han vuelto a ser tema de controversia a distintos niveles. Y, especialmente, algunos filósofos de la ciencia han tomado conciencia de que su labor no es únicamente la de aprehender la metodología formal de la investigación científica, sino que es, o debe ser, la de una reflexión crítica sobre el papel que la ciencia tiene en su interconexión con la sociedad, con el papel transformador y, en ocasiones, destructivo, de la naturaleza. Reflexión crítica que alcanza no sólo a esta labor, sino también a la propia conformación de lo que debe entenderse por ciencia. Reflexión, en cierto modo, primaria para, después, situar en su papel correspondiente las ulteriores reflexiones críticas.

Es desde este enfoque crítico desde el que cabe considerar el libro de José Sanmartín, quien ya desde el mismo título explicita su posición, su creencia en lo que debe ser la labor de un filósofo de la ciencia. Labor de no sólo pensamiento formal, pretendidamente objetivo, aséptico, sino radicalmente comprometida, desenmascaradora, a la vez, de esa pretendida asepsia, desde la que se enfoca a la ciencia como un mero conjunto de teorías sin repercusión directa sobre la realidad. Frente a la concepción heredada u ortodoxa, Sanmartín considera que toda la ciencia no es más que un proceso tecnológico que queda enmarcado en tres niveles.

El primer nivel, en el que se plasma la teoría científica, no es más que el intento de explicar los éxitos de unas técnicas previas empleadas en tradiciones operativas. Y el entramado que constituye esa interrelación entre teoría y técnica es lo que denomina tecnología. Entramado desde el cual se presenta con nitidez que la teorización de esas técnicas previas, operatorias, no hace más que localizar las causas de los sucesos para dominar su producción. Dominio con el cual conseguir, de modo inmediato, una relación de transformación de la naturaleza y, con ella, reemplazar los procesos "naturales" por otros de carácter ya "industrial". La pretendida objetividad del trabajo teórico queda, desde este enfoque, radicalmente eliminada como una creencia errónea sobreañadida a la empresa científica.

Pero toda tecnología, o teorías de primer nivel, queda arropada por teorías de un segundo nivel, el que se constituye como una teorización sobre esa tecnología. Segundo nivel desde el que se justifica el primero con cuyo ropaje pretende elaborar dicha justificación, a la vez que pretende llevarla a un plano más productivo. Cabría decir que las teorías de segundo nivel se convierten en "propaganda" de los desarrollos del futuro, de la potencia transformadora que las tecnologías de primer nivel posibilitan.

Tratando de aparecer estas teorías de segundo nivel como argumentaciones científicas —con el ropaje que de la ciencia se asume por la tradición heredada—, se apoyan, a la vez, en teorías de tercer nivel que, más que teorías, son programas de carácter metafísico, conjuntos de enunciados que, en principio, no son falsables. Más que conjunto de enunciados, proporcionan cuadros creadores de cosmovisiones, formas de captar el mundo. Así, una teoría de tercer nivel puede considerarse la formulación de la mecánica newtoniana o las teorías de la evolución darwiniana o neodarwiniana.

Con esta concepción es claro que, para Sanmartín, los científicos se escinden en dos clases: o son tecnólogos porque cultivan teorías de primer nivel, o son metafísicos porque cultivan las del tercer nivel. En alguna ocasión podrá cultivar, también, teorías de segundo nivel que, por lo dicho, por ser reflexiones sobre las del primero enmarcadas en el cuadro provisto por las del tercero, dependerán, siempre, de lo aplicado. Esto último parece una concesión, no sé si gratuita, del autor por el hecho de que algunos científicos, pocos en cantidad, ciertamente, también se hayan internado por las teorías del segundo nivel...

Ahora bien, Sanmartín no hace una exposición al modo clásico, al metodólogo objetivo como el esquema anterior que he realizado de sus ideas parecería indicar. No plasma un discurso en el mismo plano de aquellos filósofos de la ciencia a los que critica. Sus "hipótesis" las apunta, simplemente, para de modo inmediato mostrar la coherencia de las mismas haciendo camino. Y, para ello, toma como modelo el de la ingeniería genética. Ingeniería genética como tecnología de primer nivel que distingue de otras tecnologías biológicas como las de tratamiento germinal o prácticas operatorias que manejan, macroscópicamente, células germinales y cromosomas. Se limita a las tecnologías biológicas que tienen que ver con trabajos realizados dentro de los cromosomas —quitar, poner o alterar genes...—. Mostrando que es tecnología, llega a una conclusión, la de que todavía se está en un estadio más de "carpintería" que de auténtica "ingeniería"...

A pesar de esta conclusión, esta tecnología se ha visto arropada por una teoría de segundo nivel, que ha pretendido justificar la primera y que es la sociobiología. Justificar la primera y tratar de llevar su aplicación a unas prácticas que se apoyan en la idea de un cierto mundo feliz o cuadro metafísico —más allá de lo científico— en el que se mueven, realmente, los sociobiólogos. Así, si la agresividad es una constante humana y la ingeniería puede identificar "el" gen que hace agresivo al individuo humano, bastará aplicar esa ingeniería genética, esa tecnología para alterar "el" gen responsable y su-

primir, con ello, tal agresividad... Temas como el de la agresividad mencionado, la diferencia de razas, el papel de la mujer en la sociedad... son los que están presentes en este segundo nivel, sociobiológico, que pretende arroparse, en sus argumentaciones, con el rigor objetivo del trabajo científico aportado por la tecnología del primer nivel y, a la vez, la eleva para la obtención de una transferencia de la naturaleza. A su vez, es nivel que se apoya en la cosmovisión de tercer nivel, en este caso el darwinismo o más bien neodarwinismo, como base para realizar su justificación instrumental de la tecnología de primer nivel.

Haciendo camino, mostrando de facto cuál es la labor que considera propia de la reflexión crítica en torno a la ciencia, Sanmartín no puede evitar un cierto utopismo, como si él mismo también formara parte de esa labor crítica. Si el científico no es un ser aislado, sino que se encuentra inmerso en unas relaciones de poder que le llevan, no a estudiar o investigar la verdad, sino que hace conocimiento en función de una tecnología, de unas producciones apoyándose en previos procesos técnicos de carácter simplemente operatorio, a ese científico se le plantean una serie de problemas. Así, la fácil confusión de la labor científica con el negocio estrictamente lucrativo; la rendida y servicial puesta a disposición de la "defensa", bien nacional, bien de algún bloque; la consideración de que su conocimiento constituye un conocimiento excelso, sólo propio de iniciados... Y frente a estos problemas que Sanmartín estima como peligros, alza la bandera de un necesario control social y no sólo propio del elitismo de los científicos; aún más, aboga por la necesidad de una nueva moral "científica" que suprima las connotaciones mercantiles y los servilismos militares a que está sometida la tecnología científica. Desde la reflexión crítica, alcanzar la visión de una nueva ciencia entendida como empresa de conocimiento que no viole la naturaleza, que no pretenda como hasta ahora dominarla y suplirla. Utopismo de nueva moral que Sanmartín ve factible si se consigue la plasmación de un nuevo orden educativo.

Un libro que se hace breve. Y que se escinde, además, en dos grandes partes: el texto en sí y una serie de notas que tienen casi la misma extensión que el texto y que, a pesar de la advertencia del autor de que pueden dejarse a un lado en una primera lectura, se me muestran sustanciales. Un libro para meditar, para reflexionar sin dejarse arrastrar por el estilo de Sanmartín, por la apariencia de ser un libro más en torno a la moda sociobiológica, más aún a la moda biológica que algunas editoriales y medios de comunicación parecen querer imponer. Y, en esa reflexión crítica a la que este libro incita, cabrían al menos, dos preguntas: en primer lugar si la elección de ingeniería genética-sociobiológica-neodarwinismo, como modelo para hacer plausible la hipótesis subyacente, puede trasladarse a otros terrenos científicos o ese traslado exigiría modificaciones, no ya en cuanto a los niveles en sí sino en cuanto a la misma consideración de tecnología para la empresa científica en su nivel primero. Si esta consideración "tecnológica" frente a la estimada búsqueda de la verdad no depende, a su vez, de una cosmovisión que pretende acentuar, en el fondo, la vertiente pragmática de toda labor científica. Cosmovisión que se refleja en las distintas ayudas a la investigación que los distintos gobiernos establecen, y que no van, en general, sino a campos específicamente técnicos. Quiero decir, si esa visión tecnológica de lo científico no prima lo operatorio sobre lo conceptual, y colabora, en el fondo, a eliminar esto último en beneficio de los nuevos redentores...

En otro plano, el del utopismo, cabría estimar si Sanmartín, llevado por su entusiasmo crítico, por el análisis acerado pero radicalmente nítido, de una situación de la empresa científica como la plasmada en el modelo elegido —y quizá lo mismo ocurriera si el modelo hubiera sido el de la energía nuclear—, no cae en un plano en el que se le pudiera criticar el papel de "nuevo redentor" que achaca, precisamente, a los sociobiólogos. Visión de una nueva moral en la que una parte de la naturaleza, la humana, no transforme la otra con el peligro de extinción para la primera...

Obra, insisto de nuevo, para la reflexión. Y es punto en el que Sanmartín creo que logra uno de sus máximos objetivos. Por otro lado es libro que inaugura una nueva colección denominada Nueva Ciencia, con obras que intentarán seguir la línea aquí modélicamente representada de reflexión (filosofía) crítica.

Debo indicar que, en su materialidad, está bien editada, con la incorporación de fotos que, en algún caso, constituyen verdadero texto, como la que muestra las enjutas de la Catedral Nueva de Salamanca, inmersa en la crítica del enfoque que algún sociobiólogo realiza del neodarwinismo. (J.L.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

CARTOGRAFIA CROMOSOMICA CON MARCADORES DE ADN

MULTILOCUS LINKAGE ANALYSIS IN HUMANS: DETECTION OF LINKAGE AND ESTIMATION OF RECOMBINATION. G. M. Lathrop, J.-M. Lalouel, C. Julier y J. Ott en *The American Journal of Human Genetics*, vol. 37, n.º 3, págs. 482-498; mayo, 1985.

CONSTRUCTION OF HUMAN GENETIC LINKAGE MAPS, I: PROGRESS AND PERSPECTIVES. R. White, M. Leppert, P. O'Connell, Y. Nakamura, C. Julier, S. Woodward, A. Silva, R. Wolff, M. Lathrop y J.-M. Lalouel en *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 51, págs. 29-38; 1986.

INVESTIGATION OF GENETIC LINKAGE IN HUMAN FAMILIES. Ray White y Jean-Marc Laouel en *Advances in Human Genetics*, vol. 16, págs. 121-228; 1987.

VARIABLE NUMBER OF TANDEM REPEAT (VNTR) MARKERS FOR HUMAN GENE MAPPING. Yusuke Nakamura, Mark Leppert, Peter O'Connell, Roger Wolff, Tom Holm, Melanie Culver, Cindy Martin, Esther Fujimoto, Mark Hoff, Erika Kumlin y Ray White en *Science*, volumen 235, número 4796, págs. 1616-1622; 27 de mayo de 1987.

UN DEFECTO EN UN ESPEJO UNIVERSAL

EFFECTIVE HAMILTONIAN FOR $\Delta S = 1$ WEAK NONLEPTONIC DECAYS IN THE SIX-QUARK MODEL. Frederick J. Gilman y Mark B. Wise en *Physical Review D*, vol. 20, págs. 2392-2410; 1 de noviembre de 1979.

MEASUREMENT OF THE *CP*-NONCONSERVATION PARAMETER ϵ'/ϵ. J. K. Black, S. R. Blatt, M. K. Campbell, H. Kasha, M. Mannelli, M. P. Schmidt, C. B. Schwarz, R. K. Adair, R. C. Larsen, L. B. Leipuner y W. M. Morse en *Physical Review Letters*, vol. 54, n.º 15, págs. 1628-1630; 15 de abril de 1985.

THE GREAT DESIGN: PARTICLES, FIELDS, AND CREATION. Robert Adair. Oxford University Press, 1987.

EVOLUCION DEL CLIMA EN LOS PLANETAS TERRESTRES

ATMOSPHERES. Richard M. Goody y James C. G. Walker. Prentice-Hall, Inc., 1972.

CLIMATIC CHANGE ON THE TERRESTRIAL PLANETS. James B. Pollack en *Icarus*, vol. 37, n.º 3, págs. 479-553; marzo, 1979.

A NEGATIVE FEEDBACK MECHANISM FOR THE LONG-TERM STABILIZATION OF EARTH'S SURFACE TEMPERATURE. James C. G. Walker, P. B. Hays y J. F. Kasting en *Journal of Geophysical Research*, vol. 86, n.º C10, págs. 9776-9782; 20 de octubre de 1981.

LA ADAPTABLE ZARIGÜEYA

THE BIOLOGY OF SENESCENCE. A. Comfort. Elsevier North-Holland Publishing Co., 1979.

VIRGINIA OPOSSUM. Alfred L. Gardner en *Wild Mammals of North America: Biology, Management, and Economics*, dirigido por Joseph A. Chapman y George A. Feldhamer. The Johns Hopkins University Press, 1982.

SEX RATIO VARIATION IN MAMMALS. T. H. Clutton-Brock y G. R. Iason en *Quarterly Review of Biology*, vol. 61, n.º 3, págs. 339-374; 1986.

PLASTICOS QUE CONDUCEN LA ELECTRICIDAD

POLYACETYLENE: CHEMISTRY, PHYSICS, AND MATERIAL SCIENCE. James C. W. Chien. Academic Press, 1984.

HANDBOOK OF CONDUCTING POLYMERS. Terje A. Skotheim. Marcel Dekker, Inc., 1986.

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY OF SYNTHETIC METALS. En *Synthetic Metals*, vol. 17-19; enero-marzo, 1987.

RIESGO VOLCANICO

THE SURVEILLANCE AND PREDICTION OF VOLCANIC ACTIVITY. Unesco. París, 1971.

LOS VOLCANES DE LAS ISLAS CANARIAS (Volúmenes I: Tenerife, II: Lanzarote y Fuerteventura, y III: Gran Canaria). Vicente Araña y Juan Carlos Carracedo. Editorial Rueda; Madrid, 1978.

VOLCANOLOGÍA DE CANARIAS. Juan Carlos Carracedo, en "Geografía de Canarias" (capítulos 1,3,4,5 y 6). Editorial Interinsular Canaria; Tenerife, 1984.

VOLCANIC EMERGENCY MANAGEMENT. John Tomblin. Office of the United Nations Disaster Relief Co-Ordinator (UNDRO). Geneve, 1985.

THE 1985 RUIZ VOLCANO DISASTER. Darrell G. Herd y Comité de Estudios Volcanológicos; *EOS*, vol. 67, págs. 457-460; 1986.

RAMANUJAN Y EL NUMERO PI

MODULAR EQUATIONS AND APPROXIMATIONS TO π. S. Ramanujan en *The Quarterly Journal of Pure and Applied Mathematics*, vol. 45, págs. 350-372; 1914.

COMPUTATION OF π USING ARITHMETIC-GEOMETRICMEAN. E. Salaminen *Mathematics of Computation*, vol. 30, n.º 135, págs. 565-570; julio, 1976.

A HISTORY OF π. Petr Beckmann. The Golem Press, 1977.

PI AND THE AGM: A STUDY IN ANALYTIC NUMBER THEORY AND COMPUTATIONAL COMPLEXITY. Jonathan M. Borwein y Peter B. Borwein. John Wiley & Sons, Inc., 1986.

LA PESTE NEGRA

THE BLACK DEATH. Philip Ziegler. Harper & Row, Inc., 1970.

LES HOMMES ET LA PESTE EN FRANCE ET DANS LES PAYS EUROPÉENS ET MEDITERRANÉENS. Jean-Noël Biraben. Mouton et Cie., 1975.

THE BLACK DEATH IN THE MIDDLE EAST. Michael W. Dols. Princeton University Press, 1977.

JUEGOS DE ORDENADOR

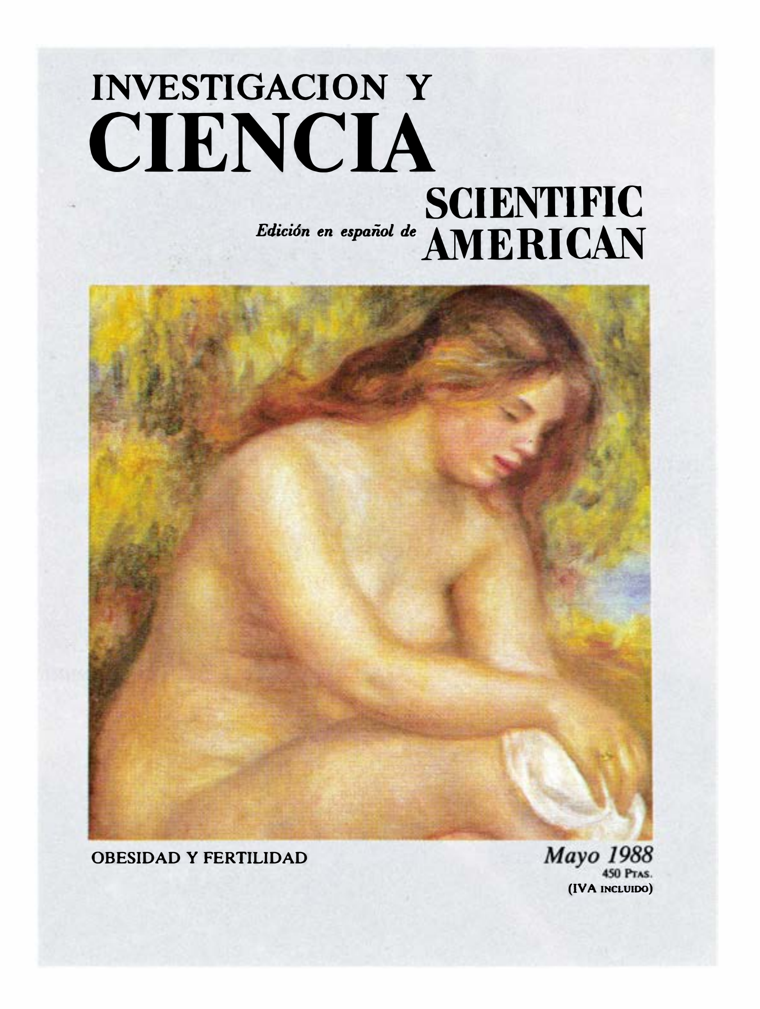
THE BLIND WATCHMAKER. R. Dawkins. W. W. Norton & Company, 1987.

TALLER Y LABORATORIO

WATER WAVES: THE MATHEMATICAL THEORY WITH APPLICATIONS. J. J. Stoker. Interscience Publishers, 1957.

ELEMENTARY DERIVATION OF THE WAKE PATTERN OF A BOAT. Frank S. Crawford en *American Journal of Physics*, vol. 52, n.º 9, págs. 782-785; septiembre, 1984.

Seguiremos explorando los campos del conocimiento



EL VALLESIENSE, por Jordi Agustí
La fauna de mamíferos del yacimiento paleontológico de Can Llobateres, en las proximidades de Barcelona, refleja un momento culminante de la diversidad animal del continente europeo hace 12 millones de años.

GRAVEDAD Y ANTIMATERIA, por Terry Goldman, Richard J. Hughes y Michael Martin Nieto
¿Caería una antimanzana a la misma velocidad que una manzana? La relatividad general y la mecánica cuántica dicen que sí, pero al hacer esa afirmación les falla la voz: ambas teorías nunca se han reconciliado.

EL PUERTO ROMANO DE COSA, por Anna Marguerite McCann
Tras dos décadas de estudio arqueológico interdisciplinario se ha retratado la economía, estructura social y nivel tecnológico de un puerto romano del Mediterráneo dominado por una emprendedora familia de nobles.

GRUAS, por Lawrence K. Shapiro y Howard I. Shapiro
La elección del modelo idóneo resulta decisiva, por razones económicas y de seguridad. En los superpoblados entornos urbanos plantean, en ocasiones, retos ingenieriles formidables.

EL DISPOSITIVO DE EFECTO CUANTICO: ¿TRANSISTOR DEL MAÑANA? por Robert T. Bate
El tamaño y el número de los elementos de un circuito susceptibles de incluirse en una pastilla de ordenador están sujetos a límites, entre otros, el comportamiento ondulatorio de los electrones, que puede afectar al funcionamiento de los dispositivos más pequeños.

LAS MANCHAS DEL LEOPARDO, por James D. Murray
En su aparición intervienen reglas matemáticas como las que rigen la vibración de las cuerdas del violín. La interacción de esas resonancias con el tamaño y la forma determinan la generación de motas, bandas u otras pautas.

OBESIDAD Y FERTILIDAD por Rose E. Frisch
La mujer debe almacenar en su organismo cierta cantidad mínima de grasa para poder reproducirse; la ganancia o la pérdida de unos pocos kilos pueden inclinar la balanza en un sentido u otro.

ANTICUERPOS CATALITICOS, por Richard A. Lerner y Alfonso Tramontano
Una nueva clase de moléculas conjuga la ingente diversidad de los anticuerpos con el poder catalítico que hace de las enzimas herramientas inestimables de la técnica, medicina y la investigación básica.

